

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-194796

(43) 公開日 平成11年(1999) 7月21日

| (51) Int.Cl. <sup>6</sup> | 識別記号  | F I           |         |
|---------------------------|-------|---------------|---------|
| G 1 0 L 3/02              |       | G 1 0 L 3/02  | A       |
| 7/04                      |       | 7/04          | C       |
| G 1 1 B 20/10             | 3 2 1 | G 1 1 B 20/10 | 3 2 1 Z |
| H 0 3 M 7/30              |       | H 0 3 M 7/30  | A       |

審査請求 有 請求項の数20 ○ L (全 29 頁)

|              |                  |          |  |
|--------------|------------------|----------|--|
| (21) 出願番号    | 特願平10-218925     | (71) 出願人 | 000003821<br>松下電器産業株式会社<br>大阪府門真市大字門真1006番地  |
| (22) 出願日     | 平成10年(1998) 8月3日 | (72) 発明者 | 三▲さき▼ 正之<br>大阪府門真市大字門真1006番地 松下電器<br>産業株式会社内 |
| (31) 優先権主張番号 | 特願平9-300121      | (72) 発明者 | 谷口 宏嗣<br>大阪府門真市大字門真1006番地 松下電器<br>産業株式会社内    |
| (32) 優先日     | 平9 (1997) 10月31日 | (72) 発明者 | 田川 潤一<br>大阪府門真市大字門真1006番地 松下電器<br>産業株式会社内    |
| (33) 優先権主張国  | 日本 (J P)         | (74) 代理人 | 弁理士 早瀬 憲一                                    |

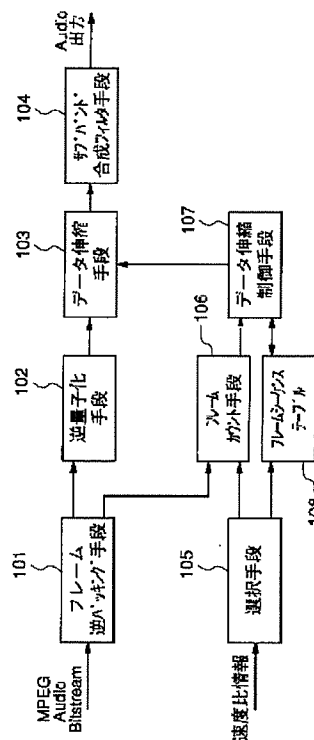
最終頁に続く

## (54) 【発明の名称】 音声再生装置

## (57) 【要約】

【課題】 フレーム単位でデコードを行う場合の速度変換処理を簡単な構成で品質良い音質を得ることを目的とする。

【解決手段】 入力される音声信号をフレーム単位で復号するフレーム逆パッキング手段101と、上記復号されたフレーム内のデータに対して時間軸変換処理を施すデータ伸縮手段103と、与えられた速度比に応じて各フレームに対する伸縮処理のシーケンスが定められたフレームシーケンステーブル108と、上記入力音声信号のフレーム数をカウントするフレームカウント手段106と、フレームカウント手段106からのカウント値をもとにフレームシーケンステーブル108を参照しデータ伸縮手段103に対して当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段107とを具備し、データ伸縮手段はデータ伸縮制御手段からの指示信号に基づいて音声信号の時間軸変換処理を行う。



**【特許請求の範囲】**

【請求項1】 入力される音声信号をフレーム単位で復号する音声復号化手段と、

上記復号されたフレーム内のデータに対して時間軸変換処理を施すデータ伸縮手段と、

与えられた速度比に応じて各フレームに対する伸縮処理のシーケンスが定められたフレームシーケンステーブルと、

上記入力音声信号のフレーム数をカウントするフレームカウント手段と、

上記フレームカウント手段からのカウント値をもとに上記フレームシーケンステーブルを参照し上記データ伸縮手段に対して当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、

上記データ伸縮手段は、上記データ伸縮制御手段からの指示信号に基づいて音声信号の時間軸変換処理を行うものである音声再生装置。

【請求項2】 請求項1に記載の音声再生装置において、

データ伸縮手段は、入力された音声信号の各フレームを少なくとも2つのセグメントに分割して各セグメントを構成する波形データを重み付け加算するクロスフェード手段を具備することを特徴とする音声再生装置。

【請求項3】 請求項1または2に記載の音声再生装置において、

データ伸縮手段は、フレームに対して所定比率の時間軸圧縮／伸長を行い、

データ伸縮制御手段は、時間軸圧縮／伸長を施すフレームと、時間軸変換なしで出力するフレームとの出現頻度を制御することにより、上記与えられた速度比で音声再生することを特徴とする音声再生装置。

【請求項4】 請求項3に記載の音声再生装置において、

データ伸縮手段は、フレームに対して所定比率の時間軸圧縮／伸長を行い、フレームシーケンステーブルには時間軸圧縮／伸長のシーケンスが繰り返されるフレームサイクル内において時間軸圧縮／伸長を施すフレームを可能な限り均等に配置することにより、上記与えられた速度比で音声を再生することを特徴とする音声再生装置。

【請求項5】 入力される音声信号をフレーム単位で復号する音声復号化手段と、

上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、

与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、

フレーム内における音声信号のエネルギーを演算するエネルギー演算手段と、

上記エネルギー演算手段の出力と、伸縮頻度制御手段の出力とをもとに、伸縮すべきフレームを選択するフレーム選択手段と、

上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、

上記フレーム選択手段は、エネルギーの小さいフレームを優先的に選択することを特徴とする音声再生装置。

【請求項6】 入力される音声信号を、フレーム単位で復号する音声復号化手段と、

上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、

与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、

当該フレームに含まれる人間の音声を含む可能性を演算する音声らしさ演算手段と、

上記音声らしさ演算手段の出力と、伸縮頻度制御手段の出力とをもとに、伸縮すべきフレームを選択するフレーム選択手段と、

上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、

上記フレーム選択手段は、音声らしさの小さいフレームを優先的に選択することを特徴とする音声再生装置。

【請求項7】 入力される音声信号をフレーム単位で復号する音声復号化手段と、

上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、

与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、

フレーム内における音声信号の定常性を演算する定常性演算手段と、

上記定常性演算手段の出力と、伸縮頻度制御手段の出力とをもとに、伸縮すべきフレームを選択するフレーム選択手段と、

上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、

上記フレーム選択手段は、定常性の高いフレームを優先的に選択することを特徴とする音声再生装置。

【請求項8】 入力される音声信号をフレーム単位で復号する音声復号化手段と、

上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、

与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、

フレーム内における音声信号のエネルギーの変化度合い

を演算するエネルギー変化度合演算手段と、  
上記エネルギー変化度合演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、  
上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、  
上記フレーム選択手段は、エネルギー変化度合をもとに、処理劣化が継時マスキング効果によって検知しにくいフレームを優先的に選択することを特徴とする音声再生装置。

【請求項9】 入力される音声信号をフレーム単位で復号する音声復号化手段と、  
上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、  
与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、  
フレーム内における音声信号のエネルギーを演算するエネルギー演算手段、あるいは、当該フレームに含まれる人間の音声を含む可能性を演算する音声らしさ演算手段、あるいは、フレーム内における音声信号の定常性を演算する定常性演算手段、あるいは、フレームにおける音声信号のエネルギー変化度合を演算するエネルギー変化度合演算手段のうち2つ以上の演算手段を備え、  
上記複数の演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、  
上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、  
上記フレーム選択手段は、上記複数の演算手段の出力を総合的に判断して、選択すべきフレームを決定することを特徴とする音声再生装置。

【請求項10】 請求項1ないし9のいずれかに記載の音声再生装置において、  
フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号を行うことを特徴とする音声再生装置。

【請求項11】 請求項1ないし10のいずれかに記載の音声再生装置において、  
フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号することを特徴とする音声再生装置。

【請求項12】 請求項5に記載の音声再生装置において、  
フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、  
エネルギー演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギーを推定することを特徴とする音声再生装置。

【請求項13】 請求項7に記載の音声再生装置において、  
フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、  
定常性演算手段は、波形定常性を示すスケールファクタ選択情報をもとに音声信号の定常性を推定することを特徴とする音声再生装置。

【請求項14】 請求項8に記載の音声再生装置において、  
フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、  
エネルギー変換度合演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギー変化度合を推定することを特徴とする音声再生装置。

【請求項15】 請求項9に記載の音声再生装置において、  
フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、  
エネルギー演算手段、定常性演算手段、エネルギー変換度合演算手段のうち2つ以上の演算手段を備え、  
エネルギー演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギーを推定し、  
定常性演算手段は、波形定常性を示すスケールファクタ選択情報をもとに音声信号の定常性を推定し、  
エネルギー変化度合演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギー変化度合を推定することを特徴とする音声再生装置。

【請求項16】 請求項1ないし15のいずれかに記載の音声再生装置において、  
データ伸縮手段は、フレームを構成する各セグメントの相関を求め、各セグメントの相関が高くなる位置を演算により求めて、当該位置にセグメントの波形データをシフトさせるためのシフト量を送出する相関演算手段を具備し、  
クロスフェード手段は、上記シフト量をもとにセグメントの波形データをシフトさせて各セグメントデータの重み付け加算を行い、  
次回に時間軸圧縮／伸長されるフレームは、前回時間軸圧縮／伸長されたフレームに関する上記シフト量を考慮のうえ、セグメントデータがシフトされ重み付け加算が施されることを特徴とする音声再生装置。

【請求項17】 請求項1ないし16のいずれかに記載

の音声再生装置において、  
データ伸縮手段は、フレームを構成する各セグメントの  
相関を求める相関演算手段を具備し、

フレーム単位で復号を行う音声復号化手段は、音声信号  
を複数の帯域に分割し当該帯域分割された各帯域毎に復  
号が成され、

上記相関演算手段は、音声信号のピッチ周波数を包含す  
る帯域のデータを用いて上記各セグメントの相関を求め  
ることを特徴とする音声再生装置。

【請求項18】 請求項16に記載の音声再生装置にお  
いて、

フレーム単位で復号を行う音声復号化手段は、音声信号  
を複数の帯域に分割し当該帯域分割された各帯域毎に復  
号が成され、

上記相関演算手段は、相関値の演算を各分割帯域毎に行  
い、最大の相関値を持つ帯域のシフト量を用いて重み付  
け加算を行うことを特徴とする音声再生装置。

【請求項19】 請求項16に記載の音声再生装置にお  
いて、

フレーム単位で復号を行う音声復号化手段は、音声信号  
を複数の帯域に分割し当該帯域分割された各帯域毎に復  
号が成され、

上記相関演算手段は、相関値の演算を各分割帯域のうち  
の最も平均エネルギーの大きい帯域に対して行うことを  
特徴とする音声再生装置。

【請求項20】 請求項1ないし4のいずれか、または  
請求項16ないし19のいずれかに記載の音声再生装置  
において、

フレームシーケンステーブルは、1つの速度比に対して  
パターンの異なる複数のシーケンステーブルを備え、  
データ伸縮手段は、予め伸縮フレームにおける各セグメ  
ントの相関値の平均値を各シーケンステーブル毎に求め、  
最も相関値の平均が高いシーケンステーブルを参照して  
処理を行うことを特徴とする音声再生装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声速度を所望の  
値に変換して聴取する事が可能な音声再生装置に関する。

【0002】

【従来の技術】音声を高効率に符号化して、記憶媒体へ  
蓄積、あるいは通信網を利用して伝送する技術が近年実  
用化され広く利用されている。

【0003】このような技術に関し、国際標準規格のMP  
EG方式を用いて、音声（オーディオ）を再生する装置と  
して、例えば特開平9-73299号公報に開示されて  
いるものがある。このMPEGオーディオ再生装置のブロッ  
ク図を図19に示す。以下、図19を参照しながら、従  
来の音声再生装置について説明する。

【0004】図19に示すように、MPEGオーディオ再生

装置1は、再生速度検出回路2、MPEGオーディオデコー  
ダ3、話速変換処理回路4、D/Aコンバータ5、オー  
ディオアンプ6から構成されている。さらに話速変換処  
理回路4は、フレームメモリ34、話速変換部35、リ  
ングメモリ32、アップダウンカウンタ33、読み出し  
クロック生成回路36で構成されている。

【0005】MPEGオーディオ再生装置1には、MPEGオー  
ディオ方式にて符号化されたMPEGオーディオストリーム  
が入力される。MPEGオーディオデコーダ3では、上記MP  
EGオーディオストリームがデジタル信号のオーディオ  
出力に復号される。MPEGオーディオの方式およびフォー  
マットの内容に関しては、現在では種々の文献等に記述  
されており、例えば「ISO/IEC IS 11172 Part3 : Audi  
o」に記載されている。

【0006】一方、例えば2倍速、0.5倍速などの速  
度情報が再生速度検出回路2に入力され、この再生速度  
検出回路にて速度情報（再生速度）を検出してデコード  
クロックを生成する。このデコードクロックは話速変換  
処理回路4およびMPEGオーディオデコーダ3へ供給され  
る。当該MPEGオーディオデコーダ3にてデコードされた  
オーディオ信号は、話速変換処理回路4に入力され、与  
えられた上記速度情報に基づき、さらに時間軸圧縮／伸  
長あるいは無音削除／挿入を施されて、所定の話速変換  
が行われ、この話速変換された出力がスピーカ23から  
再生されることとなる。

【0007】

【発明が解決しようとする課題】しかしながら、上記の  
ような、所定時間長のフレーム単位でのデコードを行う  
MPEGオーディオのような符号化方式において、複数フレ  
ーム間にまたがるデータ処理を実施する際には、多数の  
バッファメモリなどが必要かつ処理が複雑となり、ハード  
ウェア構成が大規模となる問題を生じることになる。

【0008】さらに、同様に、国際標準規格のMPEG方式  
を用いて、音声（オーディオ）を再生する装置として、  
特開平9-81189号公報に開示されているものがある。  
このMPEGオーディオ再生装置のブロック図を、図2  
0に示す。以下、図20を参照しながら、従来の音声再  
生装置について説明する。

【0009】図20に示すように、1701は、入力さ  
れる帯域信号1をTfサンプル長の1フレーム分、分割  
し保持する第1のフレーム分割装置、1702は、入力  
される帯域信号2をTfサンプル長の1フレーム分、分  
割し保持する第2のフレーム分割装置、1703は、入  
力される帯域信号3をTfサンプル長の1フレーム分、  
分割し保持する第3のフレーム分割装置、1704は、  
入力される帯域信号4をTfサンプル長の1フレーム  
分、分割し保持する第4のフレーム分割装置である。

【0010】上記において、入力される帯域信号1～4  
は、通常の時間軸信号を4帯域に帯域分割するとともに  
4分の1にダウンサンプリングするようなフィルタバン

クによって帯域分割された、それぞれの帯域信号であり、帯域信号1は、最も低域の帯域信号、帯域信号4、は最も高域の帯域信号であるとする。

【0011】1710は、音声のピッチ成分が含まれる帯域の帯域信号の前半の信号と、後半の信号とを、 $n$ サンプルだけオーバーラップさせた時の該オーバーラップ範囲における両信号間の相関値 $S(n)$ を求め、該相関値 $S(n)$ が最大値となる $n$ を $T_c$ として検出する相関関数算出装置、1711は、聴取者からの再生速度 $F$ の指定を検出する再生速度検出装置、1712は、相関関数検出範囲に制限を設けるための相関関数検出範囲制御装置、1705は、第1のフレーム分割装置1701によって分割され保持された帯域信号の前半の信号と、後半の信号とを、 $T_c$ サンプル分オーバーラップさせてクロスフェード処理する第1のクロスフェード処理装置、1706は、第2のフレーム分割装置1702によって分割され保持された帯域信号の前半の信号と、後半の信号とを、 $T_c$ サンプル分オーバーラップさせてクロスフェード処理する第2のクロスフェード処理装置、1707は、第3のフレーム分割装置1703によって分割され保持された帯域信号の前半の信号と、後半の信号とを、 $T_c$ サンプル分オーバーラップさせてクロスフェード処理する第3のクロスフェード処理装置、1708は、第4のフレーム分割装置1704によって分割され保持された帯域信号の前半の信号と、後半の信号とを、 $T_c$ サンプル分オーバーラップさせてクロスフェード処理する第4のクロスフェード処理装置、1709は、上記クロスフェード処理された4帯域の帯域信号を帯域合成する帯域合成フィルタである。

【0012】図21は、音声信号の主要ピッチ成分が含まれる周波数帯域について、その1フレーム分の時間軸波形を表した図である。図22は、図21に示した1フレームの信号を、その前半の信号部分と、後半の信号部分との2セグメントに分割して上下に並べた図である。図23は、図22における2セグメント間の相関関数を求めた値を示したグラフである。図24は、相関関数が最大となる時刻に後半の信号成分であるセグメントをずらせた様子を定性的に示した図である。図25は、2セグメント間を $T_c$ 時間オーバーラップさせてクロスフェード処理する様子を示した図である。

【0013】以上のように構成された再生装置について、以下その動作について、図21から図25を用いて説明する。まず入力される帯域信号1の1フレーム分( $T_f$ サンプル長)のデータは、図21に示すように、音声信号の主要ピッチ成分を含んでいるものとする。そして、この1フレーム分のデータは、第1のフレーム分割装置1701によって、図22に示すような同じデータ数の2セグメントに分割して保持され、第2のフレーム分割装置1702、第3のフレーム分割装置1703、第4のフレーム分割装置1704も同様に各々の帯

域信号2, 3, 4を2セグメントに分割して保持する。

【0014】そして、再生速度検出装置で得られる目標速度比 $F$ から、2セグメントをオーバーラップするデータ長である目標オーバーラップ値 $T_b$ を、次式のように求める。

$$【0015】T_b = T_f (1 - 1/F)$$

ここで、後述する位相調整を行う影響による目標速度比 $F$ からのずれを補正するための補正パラメータ $B$ (初期値は0)を考慮して、相関関数算出装置1710で、上記第1のフレーム分割装置1701の2セグメント間のオーバーラップ区間データ長が( $T_b + B$ )の前後 $m$ サンプルの範囲で相関関数を演算し、該相関関数が最大となる場合のオーバーラップ区間長 $T_c$ を求める。その結果、 $T_c$ が $T_b$ からずれることによる目標速度比からの誤差の補正を行うため、先に述べた補正パラメータ $B$ の値を以下のように更新する。

$$【0016】B \leftarrow B + T_b - T_c$$

図22は、目標速度比 $F$ が2.0の場合の、目標オーバーラップ値 $T_b (= T_f/2)$ の位置関係で、2セグメントを上下に配置した図であり、この場合に2セグメント間の相関関数を求めた結果が、図23のようになる。この例では相関が最大となるような時刻は、4であることがわかる。図24は、この相関関数の結果をもとに、2セグメント間のオーバーラップ長を $T_c$ として表した説明図である。つまり、相関関数によって、前半のセグメントに後半のセグメントの類似度合を求め、その結果、相関の高い位置までずらせると、双方のセグメントの位相が一致することになる。そのときの2セグメント間のオーバーラップ区間長が、 $T_c$ ということになる。

【0017】次に、第1のクロスフェード処理装置1705で、第1のフレーム分割装置1701によって分割され保持された2セグメントの帯域信号を、 $T_c$ 分オーバーラップさせてクロスフェード処理を行なう。同様に、第2のクロスフェード処理装置1706、第3のクロスフェード処理装置1707、第4のクロスフェード処理装置1708でも、それぞれ、第2のフレーム分割装置1702、第3のフレーム分割装置1703、第4のフレーム分割装置1704によって分割され保持された2セグメントの帯域信号を、 $T_c$ 分オーバーラップさせてクロスフェード処理を行なう。図25は、このようなクロスフェード処理の一例を示したものである。2セグメントのオーバーラップ部分に対して、互いに相補的な重みを付けた加算を行う。(a)は、前半のセグメントにフェードアウト処理した信号、(b)は、後半のセグメントにフェードイン処理した信号である。このフェードアウト処理した信号(a)と、フェードイン処理した信号(b)とを加算することにより、同図(c)のような波形となる。その後、帯域合成フィルタ1009によって、上記のようにしてクロスフェード処理された各帯域信号が帯域合成され、通常の時間軸信号が生成され

る。

【0018】以上の処理を、逐次、Tfサンプルずつの全てのフレームに関して、各帯域の信号に行うことによって、1フレーム内のデータだけで完結する高速再生が行えることとなる。

【0019】しかしながら、上記のような構成による再生装置では、次のような課題が存在する。ここでは、標準的なMPEG1 オーディオの符号化方式を例に取り、分割

帯域数を32、各帯域1フレームのデータ数を36、補正パラメータ値Bの初期値を0、基準とする相関探索幅mを4として、実際に取り得るオーバーラップ値と、相関探索する点数などを、以上に述べた従来例の方法で求め、その結果を以下の表1に示す。ここで、計算式の小数点は切り捨てて表示している。

【0020】

【表1】

| 目標速度比 F       | 1.1 | 1.3  | 1.5  | 1.7   | 1.9   | 2.0   |
|---------------|-----|------|------|-------|-------|-------|
| 目標オーバーラップ値 Tb | 3   | 6    | 12   | 14    | 17    | 18    |
| オーバーラップ値の範囲   | 0~7 | 2~10 | 8~16 | 10~18 | 13~18 | 14~18 |
| 相関探索する点数      | 8   | 9    | 9    | 9     | 6     | 5     |

【0021】まず、速度比が1.0に近い側に関して検討する。目標オーバーラップ値が小さいこともあり、オーバーラップ値の取り得る値は、かなり小さい値の範囲に留まっている。この場合の問題点として考えられるのは、クロスフェード長が短すぎることである。相関の高い位置を求めてクロスフェード処理を実施するが、クロスフェード区間を挟んだ2セグメント間の遷移期間の長さが短すぎると、セグメント中に含まれる低周波数信号は、クロスフェードによる振幅の連続性改善の効果も少なく、波形の急激な変化をもたらしてしまい、結果として不連続感の強い再生音として聴取される。このクロスフェード区間長および相関探索幅と、音質に関する評価実験は、例えば、「鈴木、三崎：電子情報通信学会音声研究会 SP90-34, 1990.8」などにPCM 音声に対する最適値を求めている。

【0022】一方、速度比が2.0に近い側に関して検討すると、目標オーバーラップ値が上限値である18に近い値であり、オーバーラップ値の上限が1セグメント長を超えることができず、相関探索点数は十分な数になっていることがわかる。また、速度比2.0の場合、オーバーラップ値を目標値である18より小さい値にとると、次回以降にこれを修正する可能性は全く無いため、目標速度を達成するには、相関による探索は行わずに固定のオーバーラップ値を取らざるを得ない。また、相関探索する点数を増加させるために、探索幅mを大きな値にすると、目標オーバーラップ値から小さい側にずれた場合の補正パラメータ値Bは正の値であり、次回の相関探索の中心とするオーバーラップ値( $T_b + B$ )の値が1セグメント長を超える( $(T_b + B) > T_f/2$ )、という不合理が生じ、速度比を修正することが困難な状況となる。そのため、探索幅mを小さな値で使用せざるを得なく、相関探索する点数が少ないため、位相の整合性が十分に改善し得ない位置でのクロスフェードを行うことになり、結果として、位相の不整合によりしわがれた声として聴取される。

【0023】このように、このアルゴリズムでは、相関関数を用いた位相の調整を行うには、不十分な状況で動作させざるを得ないため、十分な性能を出し得ない

い。さらに、以上の中間である速度比1.5近傍の比較的良好と思われる速度の範囲においても、与えられたすべてのフレームに対してクロスフェード処理を実施することになるため、処理による信号劣化が全フレームすべてに生じ、その結果、劣化の度合いが大きく感じられることになる。このように、相関関数による位相の整合性を改善する手法は、この例では十分に機能せず、かえって、目標の速度比に収束し難い、という方式上の欠点を有している。また、この例では、高速再生に対する処理を実施するのみで、低速再生に関する機能を何ら提供し得ないものである。

【0024】本発明は、上記課題に鑑み、フレーム内データで完結する、一定速度比の時間軸圧縮／伸長処理を基本とした簡素な構成によって、高品質な高速／低速の速度変換音声を実現することのできる、音声再生装置を提供することを目的とするものである。

【0025】

【課題を解決するための手段】この目的を達成するために、請求項1にかかる音声再生装置は、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して時間軸変換処理を施すデータ伸縮手段と、与えられた速度比に応じて各フレームに対する伸縮処理のシーケンスが定められたフレームシーケンステーブルと、上記入力音声信号のフレーム数をカウントするフレームカウント手段と、上記フレームカウント手段からのカウント値をもとに上記フレームシーケンステーブルを参照し上記データ伸縮手段に対して当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記データ伸縮手段は、上記データ伸縮制御手段からの指示信号に基づいて音声信号の時間軸変換処理を行うものとしたものである。

【0026】また、請求項2にかかる音声再生装置は、請求項1に記載の音声再生装置において、データ伸縮手段は、入力された音声信号の各フレームを少なくとも2つのセグメントに分割して各セグメントを構成する波形データを重み付け加算するクロスフェード手段を具備す

ることを特徴とする音声再生装置としたものである。

【0027】また、請求項3にかかる音声再生装置は、請求項1または2に記載の音声再生装置において、データ伸縮手段は、フレームに対して所定比率の時間軸圧縮／伸長を行い、データ伸縮制御手段は、時間軸圧縮／伸長を施すフレームと、時間軸変換なしで出力するフレームとの出現頻度を制御することにより、上記与えられた速度比で音声を再生することを特徴とする音声再生装置としたものである。

【0028】また、請求項4にかかる音声再生装置は、請求項3のいずれかに記載の音声再生装置において、データ伸縮手段は、フレームに対して所定比率の時間軸圧縮／伸長を行い、フレームシーケンステーブルには時間軸圧縮／伸長のシーケンスが繰り返されるフレームサイクル内において時間軸圧縮／伸長を施すフレームを可能な限り均等に配置することにより、上記与えられた速度比で音声を再生することを特徴とする音声再生装置としたものである。

【0029】また、請求項5にかかる音声再生装置は、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号のエネルギーを演算するエネルギー演算手段と、上記エネルギー演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し上記フレーム選択手段は、エネルギーの小さいフレームを優先的に選択することを特徴とする音声再生装置としたものである。

【0030】また、請求項6にかかる音声再生装置は、入力される音声信号を、フレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、当該フレームに含まれる人間の音声を含む可能性を演算する音声らしさ演算手段と、上記音声らしさ演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、音声らしさの小さいフレームを優先的に選択することを特徴とする音声再生装置としたものである。

【0031】また、請求項7にかかる音声再生装置は、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号の定常性を演算する定常性演算手段と、上記定常性演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、定常性の高いフレームを優先的に選択することを特徴とする音声再生装置としたものである。

【0032】また、請求項8にかかる音声再生装置は、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号のエネルギーの変化度合いを演算するエネルギー変化度合演算手段と、上記エネルギー変化度合演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、エネルギー変化度合をともに、処理劣化が継時マスキング効果によって検知にくいフレームを優先的に選択することを特徴とする音声再生装置としたものである。

【0033】また、請求項9にかかる音声再生装置は、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号のエネルギーを演算するエネルギー演算手段、あるいは、当該フレームに含まれる人間の音声を含む可能性を演算する音声らしさ演算手段、あるいは、フレーム内における音声信号の定常性を演算する定常性演算手段、あるいは、フレーム内における音声信号のエネルギー変化度合を演算するエネルギー変化度合演算手段のうち2つ以上の演算手段を備え、上記複数の演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段

とを具備し、上記フレーム選択手段は、上記複数の演算手段の出力を総合的に判断して、選択すべきフレームを決定することを特徴とする音声再生装置としたものである。

【0034】また、請求項10にかかる音声再生装置は、請求項1ないし9のいずれかに記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号を行うことを特徴とする音声再生装置としたものである。

【0035】また、請求項11にかかる音声再生装置は、請求項1ないし10のいずれかに記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEGオーディオ符号化方式にて符号化されたデータを復号することを特徴とする音声再生装置としたものである。

【0036】また、請求項12にかかる音声再生装置は、請求項5に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、エネルギー演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギーを推定することを特徴とする音声再生装置としたものである。

【0037】また、請求項13にかかる音声再生装置は、請求項7に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、定常性演算手段は、波形定常性を示すスケールファクタ選択情報をもとに音声信号の定常性を推定することを特徴とする音声再生装置としたものである。

【0038】また、請求項14にかかる音声再生装置は、請求項8に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、エネルギー変換度合演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギー変化度合を推定することを特徴とする音声再生装置としたものである。

【0039】また、請求項15にかかる音声再生装置は、請求項9に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、エネルギー演算手段、定常性演算手段、エネルギー変換度合演算手段のうち2つ以上の演算手段を備え、エネルギー演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギーを推定し、定常性演算手段は、波形定常性を示すスケールファクタ選択情報をもとに音声信号の定常性を推定し、エネルギー変化度合演算手段は、再生時の波形倍率を示すス

ケールファクタインデックスをもとに音声信号のエネルギー変化度合を推定することを特徴とする音声再生装置としたものである。

【0040】また、請求項16にかかる音声再生装置は、請求項1ないし15のいずれかに記載の音声再生装置において、データ伸縮手段は、フレームを構成する各セグメントの相関を求め、各セグメントの相関が高くなる位置を演算により求めて、当該位置にセグメントの波形データをシフトさせるためのシフト量を送出する相関演算手段を具備し、クロスフェード手段は、上記シフト量をもとにセグメントの波形データをシフトさせて各セグメントデータの重み付け加算を行い、次回に時間軸圧縮／伸長されるフレームは、前回時間軸圧縮／伸長されたフレームに関する上記シフト量を考慮のうえ、セグメントデータがシフトされ重み付け加算が施されることを特徴とする音声再生装置としたものである。

【0041】また、請求項17にかかる音声再生装置は、請求項1ないし16のいずれかに記載の音声再生装置において、データ伸縮手段は、フレームを構成する各セグメントの相関を求める相関演算手段を具備し、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号が成され、上記相関演算手段は、音声信号のピッチ周波数を包含する帯域のデータを用いて上記各セグメントの相関を求めることを特徴とする音声再生装置としたものである。

【0042】また、請求項18にかかる音声再生装置は、請求項16に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号が成され、上記相関演算手段は、相関値の演算を各分割帯域毎に行い、最大の相関値を持つ帯域のシフト量を用いて重み付け加算を行うことを特徴とする音声再生装置としたものである。

【0043】また、請求項19にかかる音声再生装置は、請求項16に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号が成され、上記相関演算手段は、相関値の演算を各分割帯域のうちの最も平均エネルギーの大きい帯域に対して行うことを特徴とする音声再生装置としたものである。

【0044】また、請求項20にかかる音声再生装置は、請求項1ないし19のいずれか、または請求項16ないし19のいずれかのいずれかに記載の音声再生装置において、フレームシーケンステーブルは、1つの速度比に対してパターン異なる複数のシーケンステーブルを備え、データ伸縮手段は、予め伸縮フレームにおける各セグメントの相関値の平均値を各シーケンステーブル毎に求め、最も相関値の平均が高いシーケンステーブルを参照して処理を行うことを特徴とする音声再生装置とし



たものである。

【0045】

【発明の実施の形態】（実施の形態1）以下、本発明の第1の実施の形態について、図面を参照しながら説明する。図1は本発明の第1の実施の形態における音声再生装置のブロック図を示すものである。図1において、101はフレーム逆パッキング手段、102は逆量子化手段、103はデータ伸縮手段、104はサブバンド合成フィルタ手段、105は選択手段、106はフレームカウント手段、107はデータ伸縮制御手段、108はフレームシーケンステーブルである。以下に、その動作について説明する。

【0046】本実施の形態は、MPEG1オーディオのビットストリームをデコードする際の中間データに対して速度変換処理を施す音声再生装置の例を示すものである。MPEG1オーディオのビットストリームは、ヘッダ、ビット割当て情報、スケールファクタに関する情報、サンプルデータ情報などから成り立っている。

【0047】図1において、入力されたMPEG1オーディオのビットストリームは、フレーム逆パッキング手段101によって、当該ビットストリームからヘッダ、ビット割当て情報、スケールファクタに関する情報、サンプルデータ情報などの個々の情報に分離される。逆量子化手段102では、当該逆パッキングにて得られた、各帯域（MPEG1オーディオでは32のサブバンド（帯域））毎のビット割当て情報や、スケールファクタに関連する情報をもとにして、各帯域毎に逆量子化したデータを得る。

【0048】データ伸縮手段103は、後述するデータ伸縮制御手段107からの制御によって、時間軸圧縮／伸長を施すフレームに該当する時は、逆量子化手段102の出力を一定比率で時間軸圧縮／伸長し、圧縮／伸長することなくスルーで出力するフレームに該当する場合には、逆量子化手段102の出力をそのままサブバンド合成フィルタ出力手段104へ出力する。サブバンド合成フィルタ手段104では、入力された各サブバンド（MPEG1オーディオでは32帯域）のデータが帯域合成され、当該合成により得られたオーディオ信号を出力する。

【0049】図2に、データ伸縮手段103の内部構成図を示す。同図において、2001は最も低いサブバンドに対応する逆量子化手段102の出力Q0に対して処理を施すデータ伸縮ユニット0であり、以降、2002、・・・、2032の各々は、各サブバンドの低い帯域側からの逆量子化手段102の出力Q1、・・・、Q31に対して処理を施す、データ伸縮ユニット1、・・・、データ伸縮ユニット31である。データ伸縮ユニットの内部の構成は、図2に示すように、バッファメモリ201、クロスフェード手段202、データ選択手段203で構成される。図では、データ伸縮ユニット1～データ伸縮ユ

ニット31については内部構成が記載されていないが、データ伸縮ユニット0と同一であるので、図では省略して記載している。

【0050】以下では、最も低い周波数帯域に相当する逆量子化手段102の出力データQ0に対して処理を施すデータ伸縮ユニット0の動作を示す。逆量子化手段102の出力Q0は、一旦バッファメモリ201に1フレーム分（所定時間長分）のデータだけ蓄積される。ここで、各サブバンドにおける1フレームのデータ数を、Nsとする。データ伸縮制御手段107からの制御信号により、スルーで出力するフレームに該当する場合には、データ選択手段203は、バッファメモリ201へ書き込まれているNs個のデータを、そのままサブバンド合成フィルタ手段104へ出力する。一方、データ伸縮制御手段107からの制御信号により、時間軸圧縮／伸長を施すフレームに該当する時は、クロスフェード手段202にて、バッファメモリ201内のNs個のデータを用いて、所定の圧伸比Srで時間軸圧縮／伸長を行なう。

【0051】クロスフェード手段によるクロスフェード処理、すなわち時間軸圧縮／伸長の方法を、図3を用いて説明する。図3は、時間軸圧縮／伸長を実施することで、フレームのデータ長が変化する様子を示した模式図の一例である。図3（a）は通常のフレームを示すものであり、ここでは、1フレームのデータ数Nsを、同数のデータ数（同一時間長）のセグメントであるSEG1、SEG2の2つに分割した例である。これらのセグメントを基にして、図3に示すような重み付け加算、すなわち、クロスフェード処理を行うことによって、前後の不連続無く、データ数を減少／増加させることができる。例えば、圧縮する場合は図3（b）のように行い、伸長する場合は同図（c）のようにクロスフェード処理を行う。また、クロスフェード処理を施すことなくデータ伸縮手段103にてスルーで出力するフレームである場合には、図3（a）に示すフレーム信号がそのままの状態ですばサブバンド合成フィルタ手段104へ出力される。なお同図において、（b）は圧伸比（ $=1/\text{速度比}$ ） $1/2$ で時間軸圧縮されたフレームの例、（c）は圧伸比 $3/2$ で時間軸伸長されたフレームの例である。なお、圧伸比は、 $\text{圧伸比} = 1/\text{速度比} = \text{クロスフェード手段からの出力データ数}/\text{クロスフェード手段への入力データ数}$ で定義するものとする。

【0052】図3（b）のような圧縮処理を全フレームに対して行うことにより、一定の速度比 $2/3$ の再生音を得ることができる。また、同図（c）のような伸長処理を全フレームに対して行うことにより、一定の速度比 $3/2$ の再生音を得ることができる。このような速度変換処理を行う場合には、データ伸縮制御手段107からデータ伸縮手段103へ、圧縮／伸長／スルーを示す制御信号を送り、この制御信号を基に、各データ伸縮ユニットを制御することにより、該速度変換処理を実現でき

る。例えば、上述したような速度比2.0を実現するには、入力された速度比情報(=2.0)を基に、「速度比=2.0の圧縮」を示す制御信号をデータ伸縮手段103へ出力する。クロスフェード手段202は、当該制御信号を受けとって、全フレームに対して図3(b)に示すクロスフェード処理を行い、データ選択手段203は、クロスフェード手段202の出力を選択して、サブバンド合成フィルタ手段104へ出力する。また、速度比2/3(=0.66)を実現するには、入力される速度比情報(=2/3)を基に、「速度比=2/3の伸長」を示す制御信号を、データ伸縮手段103へ出力する。クロスフェード手段202は、当該制御信号を受けとって、全フレームに対して図3(c)に示すクロスフェード処理を行い、データ選択手段203は、クロスフェード手段202の出力を選択して、サブバンド合成フィルタ手段104へ出力する。

【0053】上述した以外の速度比の音声を実現するためには、全フレームではなく特定のフレームに対してのみ、図3(b)(c)のような時間軸圧縮/伸長を行うシーケンスで処理を繰り返せば、最終的には個々のフレームの速度比とは異なった、所望の再生速度を得ることが可能となる。図4を用いてこの例を説明する。

【0054】図4は、その一例として、速度比が1.5, 1.2, 1.1, 0.9, 0.8, 0.7の場合の時間軸圧縮/伸長処理を説明するための処理シーケンス図である。同図において、(a)はスルー(時間軸圧縮/伸長処理なし)で出力するフレーム、(b)は時間軸圧縮処理を施すフレーム、(c)は時間軸伸長処理を施すフレームを

示している。また(表2)に、図4の処理速度例における、入力セグメント数、出力セグメント数、圧縮/伸長するセグメント数、繰り返すを行うフレームサイクルを示す。図4における各フレームは、図3(a)にて説明したように、同一データ数(同一時間長)の2つのセグメントから構成されており、よって、各速度比における入力・出力セグメント数、圧縮/伸長セグメント数は、(表2)の通りとなる。例えば、速度比1.5の場合を例にとると、入力セグメント数は、図4(i)の通り第1フレーム～第3フレームが入力されるので、3フレーム×2セグメント=6セグメントである。このとき、第2フレーム、第3フレームについては、時間軸圧縮処理が施されて各フレームのセグメント数が2→1になるので、圧縮セグメント数は2となり、この結果、出力セグメント数は、6セグメント-2セグメント=4セグメントとなる。速度比は(入力セグメント数/出力セグメント数)で与えられる。また(表3)に、図4に対応する、フレームシーケンステーブル108の例を示す。この例においては、テーブルには、速度比と、フレームカウント手段106でカウントするフレームサイクルと、フレームに対する圧縮/伸長/スルーの処理状態のシーケンス(フレームシーケンス)とが、記録されている。

【0055】なお、(表3)において、「a」はスルー、「b」は圧縮、「c」は伸長、を施すシーケンスを意味している。

【0056】

【表2】

| 速度比         | 0.7 | 0.8 | 0.9 | 1.1 | 1.2 | 1.5 |
|-------------|-----|-----|-----|-----|-----|-----|
| 入力セグメント数    | 14  | 4   | 18  | 22  | 6   | 6   |
| 出力セグメント数    | 20  | 5   | 20  | 20  | 5   | 4   |
| 圧縮/伸長セグメント数 | 6   | 1   | 2   | -2  | -1  | -2  |
| フレームサイクル    | 7   | 2   | 9   | 11  | 3   | 3   |

【0057】

【表3】

| 速度比 | フレームサイクル | フレームシーケンス                       |
|-----|----------|---------------------------------|
| 0.7 | 7        | a, c, c, c, c, c, c             |
| 0.8 | 2        | a, c                            |
| 0.9 | 9        | a, a, c, a, a, a, c, a, a       |
| 1.1 | 11       | a, a, b, a, a, a, a, a, b, a, a |
| 1.2 | 3        | a, b, a                         |
| 1.5 | 3        | a, b, b                         |

a: スルー  
b: 圧縮  
c: 伸長

【0058】まず、所望の速度比情報が選択手段105へ入力される。本例の場合においては、速度比=1.1, 速度比=0.7などの情報である。この速度比情報が入力されると、選択手段105は、フレームカウント手段106へフレームサイクルを、フレームシーケンステーブル108へはフレームシーケンスを送出する。この際に送出されるフレームサイクル、フレームシーケンスは、

(表3)に示されるような値である。

【0059】以下、再生時間を短くする(速度比>1.0; 時間軸圧縮処理)例を、速度比1.1の場合を例にとって説明する。

【0060】速度比情報1.1が選択手段105へ入力されると、選択手段105はフレームカウント手段106へフレームサイクル「11」を、またフレームシーケ

ステابل 1 0 8 へフレームシーケンス「a, a, b, a, a, a, a, a, b, a, a」を送出する。このフレームシーケンスは、フレームシーケンスステابل 1 0 8 に書き込まれる。フレームカウント手段 1 0 6 は、選択手段 1 0 5 からフレームサイクル「1 1」を受け取ったタイミング以降に、フレーム逆パッキング手段 1 0 1 から出力されフレームカウント手段 1 0 6 へ入力されたフレームをカウントし、フレームカウント値を出力する。この際、フレームカウント手段 1 0 6 のカウント値は、1 → 2 → … → 1 0 → 1 1 → 1 → … と、1 1 サイクルでカウントされるものとする。

【0 0 6 1】データ伸縮制御手段 1 0 7 は、当該カウント値をもとに、まず、カウント値「1」が入力されたときはフレームシーケンスステابل 1 0 8 からフレームシーケンス 1 番目のシーケンス「a」を読み込み、データ伸縮手段 1 0 3 へ「スルー」を指示する制御信号を出力する。データ伸縮手段 1 0 3 において、当該手段内部の各データ選択手段は、この「スルー」を意味する制御信号を基に、逆量子化手段 1 0 2 から出力された Q0, Q1, …, Q31 なるデータを、スルー（クロスフェード処理なし）で出力する（C0, C1, …, C31）。サブバンド合成フィルタ手段 1 0 4 では、当該 3 2 帯域の C0, C1, …, C31 を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【0 0 6 2】次に、フレームカウント手段 1 0 6 からカウント値「2」が出力されると、データ伸縮制御手段 1 0 7 は、フレームシーケンスステابل 1 0 8 からフレームシーケンス 2 番目のシーケンス「a」を読み込み、データ伸縮手段 1 0 3 へ「スルー」を指示する制御信号を出力する。これ以降の処理は、上述したカウント値「1」の場合と同じである。なお、図 4、（表 2）からも明らかなように、カウント値「4」「5」「6」「7」「8」「1 0」「1 1」の場合にフレームシーケンスステابل 1 0 8 から読み込まれるシーケンスは「a」であり、この際の処理は上述したカウント値「1」の場合と同一なので、説明を省略する。

【0 0 6 3】図 4 および（表 3）より、フレームカウント値が「3」、「9」の場合には、フレームシーケンスステابل 1 0 8 からは、フレームシーケンスとして「b」が読み込まれ、時間軸圧縮処理が施される。これについて、以下に説明する。

【0 0 6 4】フレームカウント手段 1 0 6 からカウント値「3」、「9」が出力された場合には、データ伸縮制御手段 1 0 7 は、フレームシーケンスステابل 1 0 8 からフレームシーケンス「b」を読み込み、これによりデータ伸縮手段 1 0 3 へ「圧縮」を指示する制御信号を出力する。データ伸縮手段 1 0 3 において、当該手段内部の各データ選択手段は、この「圧縮」を意味する制御信号を基に、データ伸縮ユニット 0 ～データ伸縮ユニット 3 1 内の各クロスフェード手段にて、図 3（b）を用い

て上記説明した時間軸圧縮処理を行い、この圧縮処理が施された信号がデータ選択手段 2 0 3 にて選択されて、出力される（C0, C1, …, C31）。サブバンド合成フィルタ手段 1 0 4 では、当該 3 2 帯域の C0, C1, …, C31 を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【0 0 6 5】上述のような処理にて各フレームに対して、スルー、時間軸伸長の処理が成され、フレームシーケンス「1 1」で 1 サイクルの処理が行われる。この 1 サイクル処理が終わると、その後入力されてくる各フレームに対して、上述したシーケンスと同一のシーケンスで処理が継続される。

【0 0 6 6】次に、再生速度を遅くする（速度比 < 1.0；時間軸伸長処理）例を、速度比 0.7 の場合を例にとりて説明する。速度比情報 0.7 が選択手段 1 0 5 へ入力されると、選択手段 1 0 5 はフレームカウント手段 1 0 6 へフレームサイクル「7」を、またフレームシーケンスステابل 1 0 8 へフレームシーケンス「a, c, c, c, c, c, c」を送出する。このフレームシーケンスは、フレームシーケンスステابل 1 0 8 に書き込まれる。フレームカウント手段 1 0 6 は、選択手段 1 0 6 からフレームサイクル「7」を受け取ったタイミング以降に、フレーム逆パッキング手段 1 0 1 から出力されフレームカウント手段 1 0 6 へ入力されたフレームをカウントし、フレームカウント値を出力する。この際、フレームカウント手段 1 0 6 のカウント値は、1 → 2 → … → 6 → 7 → 1 → … と、7 サイクルでカウントされるものとする。

【0 0 6 7】データ伸縮制御手段 1 0 7 は、当該カウント値をもとに、まず、カウント値「1」が入力されたときは、フレームシーケンスステابل 1 0 8 からフレームシーケンス 1 番目のシーケンス「a」を読み込み、データ伸縮手段 1 0 3 へ「スルー」を指示する制御信号を出力する。データ伸縮手段 1 0 3 において、当該手段内部の各データ選択手段は、この「スルー」を意味する制御信号を基に、逆量子化手段 1 0 2 から出力された Q0, Q1, …, Q31 なるデータを、スルー（クロスフェード処理なし）で出力する（C0, C1, …, C31）。サブバンド合成フィルタ手段 1 0 4 では、当該 3 2 帯域の C0, C1, …, C31 を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【0 0 6 8】次に、フレームカウント手段 1 0 6 からカウント値「2」が出力されると、データ伸縮制御手段 1 0 7 は、フレームシーケンスステابل 1 0 8 からフレームシーケンスとして「c」が読み込まれ、時間軸伸長処理が施される。これについて、以下に説明する。

【0 0 6 9】フレームカウント手段 1 0 6 からカウント値「2」が出力された場合には、データ伸縮制御手段 1 0 7 は、フレームシーケンスステابل 1 0 8 からフレームシーケンス「c」を読み込み、これによりデータ伸縮

手段103へ「伸長」を指示する制御信号を出力する。データ伸縮手段103において、当該手段内部の各データ選択手段は、この「伸長」を意味する制御信号を基に、データ伸縮ユニット0～データ伸縮ユニット31内の各クロスフェード手段にて、図3(c)を用いて上記説明した時間軸伸長処理を行い、この伸長処理が施された信号がデータ選択手段203にて選択されて、出力される(C0, C1, ..., C31)。サブバンド合成フィルタ手段104では、当該32帯域のC0, C1, ..., C31を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【0070】次に、フレームカウント手段106からカウント値「3」が出力されるが、図4および(表3)からも明らかなように、カウント値「3」「4」「5」「6」「7」の場合にフレームシーケンステーブル108から読み込まれるシーケンスは、第2フレームと同様に「c」であり、この際の処理は上述したカウント値「2」の場合と同一なので、説明を省略する。

【0071】上述のような処理にて、各フレームに対して、スルー、時間軸伸長の処理が成され、フレームシーケンス「7」で1サイクルの処理が行われる。この1サイクル処理が終わると、その後入力されてくる各フレームに対して、上述したシーケンスと同一のシーケンスで処理が継続される。

【0072】以上の説明より明らかなように、フレームサイクルで所望の速度比のデータ数(セグメント数)になるように、時間軸圧縮/伸長するフレームを偏りがあまりないように挿入することにより、特定のフレームサイクル内で所望の速度比を得ることが可能となる。また図4、(表2)(表3)の例とは異なる速度比の場合でも、速度比に合うように、時間軸圧縮/伸長するフレームを偏りがあまりないように挿入するシーケンステーブルを用いてフレームサイクルを繰り返すことにより、所望の速度比の音声を得ることが可能である。また、図4、(表2)(表3)の例とは異なる順番であるシーケンスパターンの場合でも、(表2)に示すような圧縮/伸長セグメント数が守られておれば、所望の速度比が得られる。

【0073】このように、一定値(本実施形態では図3のように圧縮比は $1/2$ 、伸長比は $3/2$ )の時間軸圧縮/伸長を行うフレームを所定の順番で実施するように制御すれば、所望の速度比の音声を得ることが可能となる。

【0074】なお、以上の説明においては、図3に示したように、基準とする時間軸圧縮比の値を $1/2$ 、時間軸伸長比の値を $3/2$ で実現した例で説明したが、その他の時間軸圧縮比/伸長比をもとにシーケンステーブルを構成することも、同様に実施可能である。

【0075】(実施の形態2)以下、本発明の第2の実施の形態について、図面を参照しながら説明する。第2

の実施の形態における音声再生装置の構成図は、上述した第1の実施の形態の構成図(図1)と基本的に同様の構成であり、MPEG1オーディオストリームを入力する例である。フレーム逆パッキング手段101、逆量子化手段102、サブバンド合成フィルタ手段104、選択手段105、フレームカウント手段106、フレームシーケンステーブル108、データ伸縮制御手段107、は第1の実施の形態と同様の動作をするものである。本第2の実施の形態が第1の実施の形態と異なっている点は、データ伸縮手段103の内部の構成および動作にある。

【0076】本第2の実施の形態におけるデータ伸縮手段の構成図を、図5に示す。

【0077】同図において、2001は最も低いサブバンドに対応する逆量子化手段102の出力Q0に対して処理を施すデータ伸縮ユニット0であり、以降、2002, ..., 2032の各々は、各サブバンドの低い帯域側からの逆量子化手段102の出力Q1, ..., Q31に対して処理を施す、データ伸縮ユニット1, ..., データ伸縮ユニット31である。各データ伸縮ユニットの内部の構成は、図5に示すように、バッファメモリ201、クロスフェード手段202、データ選択手段203で構成される。図ではデータ伸縮ユニット1～データ伸縮ユニット31については内部構成が記載されていないが、データ伸縮ユニット0と同一であるので、図では省略している。本実施形態の構成は、さらに、図5に示すように、相関演算手段301、位相制御記憶手段302を付加した構成となっている。

【0078】以下に、相関演算手段301と、位相制御記憶手段302の動作を中心に説明を行う。第1の実施形態においては、時間軸波形のクロスフェードは一意に定位置で重み付け加算されている。この場合、波形の振幅に関しては不連続無く接続されるが、位相に関しては考慮されていない。そこで、本実施形態においては、位相の整合性が高い位置を相関関数を用いて求め、その位置にシフトしてから重み付け加算を行うようなクロスフェード処理を行うようにする。図6に、このような重み付け加算を行ったクロスフェード処理(圧縮)の例を示す。図6(a)は、図3(a)に相当する、クロスフェード処理を施す前の元のフレームを示しており、同一データ数のセグメント1と、セグメント2とから成っている。図6(b)は、セグメント1と、セグメント2が、相関を考慮したシフトが成されることなく重み付け加算されており、これを図3(b)の圧縮フレームと同一の基準形と考える。図6(c)は、相関の高い位置が基本形の場合に比べて右に存在した場合のクロスフェード処理後のフレームであり、クロスフェード区間は、同図(b)の基準形に比べて短くなるとともに、データ量は、同図(b)に比べて増加する。逆に図6(d)は、相関の高い位置が左に存在した場合のクロスフェード処

理後のフレームであり、クロスフェード区間は、同図(b)の基準形に比べて短くなるとともに、データ量は同図(b)に比べて減少する。

【0079】上述の如き、位相の整合性を改善する目的のために、相関関数を用いたクロスフェード処理を行う速度変換装置については、本願出願人により種々の提案が成されており、例えば、本願出願人の先願たる特開平4-104200号公報(特許登録2532731号)などに示される通りである。本実施形態では、このような相関関数を用いたクロスフェード手法を用いるが、この際図5において、最も低域のデータであるQ0には、音声のピッチ周波数が存在する範囲が含まれると考えられるので、このピッチ周波数に相当する成分に関して位相の整合性を改善するために、Q0に相当する帯域データのみを用いて、相関演算手段301、位相制御記憶手段302により相関演算を行う。相関演算を行うデータは、バッファメモリ201に存在しているが、相関演算の範囲は、上記した特開平4-104200号公報などに示されるように、与えられたフレームシーケンスの値が、圧縮フレームか伸長フレームのいずれであるかと、前回求めた相関シフト量とによって決定される。

【0080】図6(c)(d)からもわかるように、相関の高い位置にシフトした場合には、本来目標としているデータ数(図6(b))に比較して過不足を生じることになる。その過不足の値は、相関の高い位置にシフト(相関シフト量を $rk$ とする)したデータ量から求めることができ、これを次回生じる時間軸圧縮/伸長処理の際に補うことにする。そのためには、データの過不足に相当する相関シフト量 $rk$ を、一旦、位相制御記憶手段302に記憶する必要がある。このシフト量 $rk$ は、次のクロスフェード処理を行う際の、加算する先頭データの位置(ポインタ)を調整することにより、補正できることになる。

【0081】このようなシフト量 $rk$ の補正を行う様子を、図7に模式的に示す。以前の圧縮フレームにおいて、基準形(図7(a))のようにシフトが生じなかった場合、ポインタP2の位置のシフトは無く、図のような位置関係で相関の高い位置を探索するので、今回の基準形でも、セグメント1と、セグメント2は、シフト無くクロスフェード処理される。以前の圧縮フレームにおいて、正方向にシフト( $rk > 0$ )した位置で重み付け加算が行われた場合(図7(b))、以前に余分にデータを出力しているので、今回のポインタ位置はP2が正方向にシフトした位置となり、今回の基準形では、セグメント1内の後ろ部分と、セグメント2内の前部分とが使用されないことになり、よってこの際の基準形は、図7(b)の如くなる。また、以前の圧縮フレームにおいて、負方向にシフト( $rk < 0$ )した位置で重み付け加算が行われた場合(図7(c))、以前にデータを不足させているので、今回のポインタ位置は、P2が負方向にシ

フトした位置となり、今回の基準形では、セグメント1内の後ろ部分は複数回(この場合2回)使用されることになり、よってこの際の基準形は、図7(c)の如くなる。いずれの場合でも、図7に示すような処理を施すことによって、今回の伸縮フレームにおける基準形の圧縮が行われる時には、以前のフレームのデータ量の、目標とするデータ量に対する誤差は吸収されていることになり、よって誤差の累積は無いことになる。上述の例では、圧縮フレームに関して説明を行ったが、伸長フレームに関して同様の考え方で実現できることは言うまでもない。このように、以前の圧縮/伸長フレームのシフト量を考慮して、ポインタ位置をシフトした位置を基準として、相関関数で相関の高い位置を求めることになる。

【0082】以上のように求められた相関シフト量 $rk$ は、他のサブバンドにおいても同様に適用してクロスフェード処理が行われ、Q0に対するクロスフェード処理と同様の処理がQ1~Q31に対しても行われる。これにより、各サブバンドにて、同一のシフト量 $rk$ にてクロスフェード処理が施されたのち、C0~C31の出力信号が合成されることになる。

【0083】以上のように、本実施の形態2の構成によれば、相関演算手段301によって位相の整合性の高い位置で重み付け加算を行うクロスフェード処理を行うことで、データ伸縮手段103の出力信号の振幅・位相の両方が、前後のフレームに対して不連続無く接続されるため、音質の向上を達成することができる。

【0084】なお、上記実施の形態2では、最低域のサブバンドの逆量子化出力データQ0に対して相関関数を求めており、音声に対する基本周波数を元に、位相の整合性を改善することに主眼をおいているが、MPEG符号化などの音声(スピーチ)信号以外の音源の場合には、必ずしも、最低域のサブバンドについて相関関数を求めることが良い結果をもたらすとは限らない。そのため、各サブバンドの逆量子化手段の出力データのすべて(第1、第2の実施形態の例でいうなら、Q0~Q31)に対して相関の高い位置を求め、その各サブバンドの最大相関値の中で最も大きいサブバンドの相関値を元に、重み付け加算するシフト量を決定することにより、周期性の高い帯域を中心とした位相の整合性を改善させることが可能となる。また、各サブバンドの平均エネルギーを求め、その最も平均エネルギーの大きいサブバンドに対して相関の高い位置を求めることによっても、同様の改善を達成することができる。

【0085】さらに、本実施の形態2の説明で述べたような1つの速度比に対して1つのフレームシーケンスを用いる1対1対応でなく、例えば図8に示すように、1つの速度比に対して伸縮フレームの発生位置が異なる複数のフレームシーケンステーブルを用意しておき(図8の例は速度比が1.1の場合)、伸縮フレームにおける相

閾値の平均を、各フレームシーケンステーブル毎に予め求めて、最も相関値の平均が高いシーケンステーブルを参照して伸縮処理を実施するようにして、伸縮フレームを発生させる位置を、より最適な位置のもので行うことにより、位相の整合性の改善度を高めることが可能となる。さらに、先に述べた各帯域における相関値の中で最大相関値を採用する方法と組み合わせれば、一層よい改善を発揮することができる。

【0086】（実施の形態3）以下、本発明の第3の実施の形態について、図面を参照しながら説明する。図9は本発明の第3の実施の形態による音声再生装置のブロック図を示すものである。図9において、3001はフレーム復号化手段、3002はデータ伸縮手段、3003は伸縮頻度制御手段、3004はエネルギー演算手段、3005はフレーム選択手段、3006はデータ伸縮制御手段である。以下に、その動作について説明する。

【0087】本実施の形態3は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。

【0088】図9において、最初に、伸縮頻度制御手段3003は、与えられた速度比の情報をもとに、速度変換処理の一連の処理の1周期に相当するフレームサイクル数 $N_f$ と、そのフレームサイクル数内で伸縮処理を行うフレーム数 $N_s$ とを出力する。そして、エネルギー演算手段3004では、伸縮頻度制御手段で決定されたフレームサイクル数分の音声のエネルギーを求める。次に、フレーム選択手段3005は、先に求められた $N_f$ 個のエネルギーの値を参考に、音声が存在しない無音状態のフレームはエネルギーが小さく、そのフレームを伸縮処理しても劣化は検知され難いと仮定し、速度変換処理のために伸縮すべきフレームとして、エネルギーの小さいフレームから優先的に所定数 $N_s$ 個の選択を行う。そして、データ伸縮制御手段3006は、当該フレームが伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段3002が、伸縮処理をすべきかどうかを制御する。その結果、入力された符号化データは、フレーム復号化手段3001で1フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて、波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめエネルギー演算手段で求めた音声のエネルギーを用いて、フレーム選択手段で、フレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難くするようにする、ことが可能となる。

【0089】なお、本実施の形態3では、各エネルギーの値を参考に、音声が存在しない無音状態のフレームはエネルギーが小さいと仮定し、伸縮すべきフレームとし

て、エネルギーの小さいフレームから優先的に所定数のフレームを選択するようにしているが、各フレームにおける平均振幅の値を用いる場合にも、有効であると考えられる。

【0090】（実施の形態4）以下、本発明の第4の実施の形態について、図面を参照しながら説明する。図10は本発明の第4の実施の形態による音声再生装置のブロック図を示すものである。図10において、3001はフレーム復号化手段、3002はデータ伸縮手段、3003は伸縮頻度制御手段、4004は音声らしさ演算手段、4005はフレーム選択手段、3006はデータ伸縮制御手段である。以下に、その動作について説明する。

【0091】本実施の形態4は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図10において、フレーム復号化手段3001、データ伸縮手段3002、伸縮頻度制御手段3003、データ伸縮制御手段3006は、実施の形態3と同様の動作を行うものである。本実施の形態4では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段の働きを、中心に説明を行う。

【0092】ここでは音声らしさという尺度をもとに、選択すべきフレームの判定を行う。ここで、音声らしさに関して説明を行う。実環境などにおける、通信や放送などでの音声信号においては、全くの無音状態あるいはそれに近い状態という状況は、ほとんどありえない。必ず背景騒音や目的としない音が混入し、目的とする音声信号に重畳する形で含まれている。つまり、より厳密に人間の音声を含むフレームを選択するには、エネルギーの大小だけではなく、含まれるフレームの性質を別の観点で分析する必要がある。そこで、該当するフレームにどのくらいの確からしさで音声信号が含まれているかを推定する尺度として、「音声らしさ」の定義を示す。中藤らによる、「ファジー推論による音声／雑音判別手法の検討」（1993年電子情報通信学会春季大会、A-223）による手法で、母音・無声摩擦音の発生頻度をファジー推論することにより、会話の音声らしさを求めて、これと予め求めてある閾値との比較によって、入力信号が、音声／雑音のいずれであるかの2者択一の判定を行っている。この音声らしさは、特定の時間内に音声が含まれる可能性を示す尺度として用いれば、雑音と音声の混入している音声でも、最も音声が含まれないと予想されるフレームを推定することができる。また、音声らしさの度合を数値化していることにより、複数フレームの音声らしさの大小に基づく相対比較判定に利用することができる。

【0093】人間が自然に発声する音声を速度別に分析すると、該人間が自然に発生する音声は、言語情報を担う音声区間以外の発声器官が休止しているポーズ区間長を伸縮させている度合いが大きいことが判っている（参考

文献2)参照)。従って、自然な音声速度変換処理を実施するためには、ポーズ区間であるところの非音声区間を伸縮の方が好ましい。

【0094】音声らしさ演算手段4004では、伸縮頻度制御手段で決定されたフレームサイクル数分の音声らしさを求める。次に、フレーム選択手段4005は、先に求められたNf個の音声らしさの値を参考に、音声らしさが小さいフレームは音声情報が少なく、そのフレームを伸縮処理しても劣化は検知され難いと仮定し、速度変換処理のために伸縮すべきフレームとして、音声らしさの小さいフレームから優先的に所定数Ns個の選択を行う。そして、データ伸縮制御手段3006は、伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段3002が伸縮処理をすべきかどうかを制御する。その結果、入力された符号化データはフレーム復号化手段3001で1フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめ、音声らしさ演算手段で求めた音声のエネルギーを用いて、フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難いものとする事が可能となる。

【0095】(実施の形態5)以下、本発明の第5の実施の形態について、図面を参照しながら説明する。図11は本発明の第5の実施の形態による音声再生装置のブロック図を示すものである。図11において、3001はフレーム復号化手段、3002はデータ伸縮手段、3003は伸縮頻度制御手段、5004は定常性演算手段、5005はフレーム選択手段、3006はデータ伸縮制御手段である。以下に、その動作について説明する。

【0096】本実施の形態5は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図11において、フレーム復号化手段3001、データ伸縮手段3002、伸縮頻度制御手段3003、データ伸縮制御手段3006は、実施の形態3と同様の動作を行うものである。本実施の形態5では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段の働きを中心に説明を行う。

【0097】本実施の形態5では、音声波形の定常性に着目する。ここでは、フレーム内における正規化自己相関関数を求め、その値の大きいものほど定常性が高いと考える。これは、時間伸縮処理は時間軸波形の類似区間をもとに波形の挿入・間引き操作を行う場合、相関の高いフレームでは波形の重み付け加算による伸縮処理を行うため、処理劣化が検知され難い定常性の高いフレームを選択して、伸縮処理を行うことにする。逆に、音声の子音の始端部分などの非定常な過渡的な部分では、重み

付け加算による劣化が顕著となる。

【0098】定常性演算手段5004では、伸縮頻度制御手段3003で決定されたフレームサイクル数分の定常性を、予め求める。次に、フレーム選択手段5005は、先に求められたNf個の定常性の値を参考に、定常性が大きいフレームは波形の周期性が高く波形の類似性が高いため、そのフレームを伸縮処理しても劣化は検知され難いと仮定し、速度変換処理のために伸縮すべきフレームとして、定常性の大きいフレームから優先的に所定数Ns個の選択を行う。そして、データ伸縮制御手段3006は、伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段3002が伸縮すべきかどうかを制御する。その結果、入力された符号化データは、フレーム復号化手段3001で1フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて、波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめ定常性演算手段で求めた音声の定常性を用いて、フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難いものとする事が可能となる。

【0099】なお、本実施の形態5では、各フレームにおける定常性を示す値として、正規化自己相関関数を利用しているが、例えば、周波数スペクトルの変化度合などを用いることも有効であると考えられる。

【0100】(実施の形態6)以下、本発明の第6の実施の形態について、図面を参照しながら説明する。図12は本発明の第6の実施の形態における音声再生装置のブロック図を示すものである。図12において、3001はフレーム復号化手段、3002はデータ伸縮手段、3003は伸縮頻度制御手段、6004はエネルギー変化度合演算手段、6005はフレーム選択手段、3006はデータ伸縮制御手段である。以下に、その動作について説明する。

【0101】本実施の形態6は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図12において、フレーム復号化手段3001、データ伸縮手段3002、伸縮頻度制御手段3003、データ伸縮制御手段3006は、実施の形態3と同様の動作を行うものである。本実施の形態3では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段の働きを中心に説明を行う。

【0102】本実施の形態6では、音声波形のエネルギー変化度合に着目する。ここでは、1フレーム内をさらに複数の小区間に分割した各小区間でのエネルギー値を求め、各小区間の前値との差分値を求めることにより、エネルギーの変化度合を求める。そして、このエネルギーの時間的な変化度合を継続的に監視することによって、時間的に継続する区間に対するマスキング効果であ

る、継時マスクング (temporal masking) の影響を考慮した処理フレームの選択を行う。このマスクングに関しては、参考文献1: Moore の本、に詳しく記述されているが、マスクの前後の双方の区間に対してマスクング効果を生じ、この性質を利用すれば、時間伸縮処理による劣化を検知され難くできる。すなわち、大きなエネルギーのフレームの直後の小さなエネルギーのフレームは、マスク (backward masking) され、時間軸伸縮の劣化が検知され難い。あるいは、小さなエネルギーのフレームに継続して直後に大きなエネルギーのフレームが到来する場合、前の小さいエネルギーのフレームは、マスク (forward masking) され、時間伸縮処理の劣化は、検知されにくい。また、これらのマスクング量は、マスクとのレベル差、および時間差によって値が異なっている。ただし、高速再生時における時間軸圧縮処理により新たに発生する継時マスクング効果によって、新たにエネルギーの小さい部分の聴き取りが困難になる、ということがないように注意する必要がある。

【0103】エネルギー変化度合演算手段6004では、伸縮頻度制御手段で決定されたフレームサイクル数分のエネルギー変化度合を予め求める。次に、フレーム選択手段6005は、先に求められたNf個のエネルギー変化度合の値を参考に、継時マスクング効果による処理劣化が検知されにくいフレームから優先的に所定数Ns個の選択を行う。その際、時間軸圧縮を行うことにより、エネルギーの小さい区間の聴き取りが困難になる、ということがないように注意しなければならない。すなわち、エネルギーの大きいフレームに挟まれたエネルギーの小さいフレームは、時間長が短くなることによる、前方・後方マスクング効果の増大が予想されるため、ほかのフレームを選択するようにする。そして、データ伸縮制御手段3006は、当該フレームが伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段3002が伸縮すべきかどうかを制御する。その結果、入力された符号化データは、フレーム復号化手段3001で1フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて、波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめエネルギー変化度合演算手段で求めたエネルギーの変化度合を用いて、フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知されにくいものとするのが可能となる。

【0104】なお、本実施の形態6では、各フレームにおけるエネルギー変化度合を示す値を指標として継時マスクング効果を利用しているが、例えば1フレーム内をさらに複数の小区間に分割した各小区間ごとの平均振幅値を求め、各小区間の前値との差分値を求めることにより、平均振幅の変化度合を代用して用いることも有効で

あると考えられる。

【0105】(実施の形態7) 以下、本発明の第7の実施の形態について、図面を参照しながら説明する。図13は本発明の第7の実施の形態による音声再生装置のブロック図を示すものである。図13において、3001はフレーム復号化手段、3002はデータ伸縮手段、3003は伸縮頻度制御手段、4004は音声らしさ演算手段、5004は定常性演算手段、6004はエネルギー変化度合演算手段、7005はフレーム選択手段、3006はデータ伸縮制御手段である。以下に、その動作について説明する。

【0106】本実施の形態7は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図13において、フレーム復号化手段3001、データ伸縮手段3002、伸縮頻度制御手段3003、データ伸縮制御手段3006は、実施の形態3と同様の動作を行うものである。また、音声らしさ演算手段4004は、実施の形態4と、定常性演算手段5004は、実施の形態5と、エネルギー変化度合演算手段6004は、実施の形態6と同様の動作を行う。本実施の形態7では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段7005の働きを中心に説明を行う。

【0107】速度変換処理によって処理された音声から得るべき情報は、音声言語情報であると仮定すると、対象とする音声処理により加工されたことによって、聴取者の了解性が低下することは望ましくない。あるいは、速度変換処理を適用することによって了解性を高めることができる可能性があることは、学会発表等より明らかにされつつある(参考文献3, 4)。例えば、音声聴取の際の時間処理能力が低下している高齢者においては、速度を低下させることによって、了解性が高められることが確認されている。本実施の形態7では、速度変換処理によって了解性を向上させ、処理による劣化を最小限に抑える、あるいは、速度変換処理によって自然性が劣化せず効率的に音声情報を聴取しやすくする、の2つの処理形態を提供するものである。フレーム選択手段7005は、音声らしさ演算手段の出力結果と、定常性演算手段の出力結果と、エネルギー変化度合演算手段によって得られるマスクング条件とをともに、各フレームに対する分析結果を数値化し、これをもとに、自然性を重視した場合、了解性を重視した場合、の双方に関して、選択すべきフレームを決定するものである。

【0108】まず、自然性の劣化を少なく、効率的に聴取する場合の処理を説明する。この場合は、音声らしさ演算手段によって得られた非音声区間のフレームに対する優先度を大きくする。そして、残りの2つの分析結果を考慮して、最終的なフレーム選択を決定する。

【0109】次に、了解性を高め、聴き取りやすい音声を得る場合の処理を説明する。この場合は、エネルギー



の小さい子音語頭部が継時マスキングされないようにエネルギー変化度合のパラメータの優先度を高くする。そして、残りの2つの分析結果を考慮して、最終的なフレーム選択を決定する。

【0110】このように、あらかじめエネルギー変化度合演算手段で求めたエネルギーの変化度合を用いて、(あるいは、音声らしさ演算手段によって得られた非音声区間のフレームに対する優先度を大きくして、)フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、自然性・了解性の優先度合いを考慮した波形の伸縮を行うことができるものである。

【0111】なお、本実施の形態7では、請求項9に対応するうちの一例として、エネルギー演算手段、音声らしさ演算手段、定常性演算手段、エネルギー変化度合演算手段の4つの手段のうち、後者の3つを備えたものを説明したが、エネルギー演算手段を判定条件に加えてどのフレームに対して時間軸伸縮を加えるべきかを、より厳密に推定することも可能である。本発明では、これは4つの演算手段のうち2つ以上を備えて総合的な推定を行うことで、再生音の聴取条件などに関して複数の選択肢を与えるものである。

【0112】(実施の形態8)以下、本発明の実施の形態8について、図面を参照しながら説明する。まず、以下の実施の形態8～11の説明に先立ち、MPEG1オーディオレイヤ1/2符号化方式について説明する。MPEG1オーディオレイヤ1/2符号化方式は、図26に示すブロック図で表される。16ビット直線量子化された入力信号は、サブバンド分析フィルタで32帯域のサブバンド信号に分割される。フィルタは、512タップPFB (Polyphase Filter Bank) で実現される。各サブバンド信号に対してスケールファクタを計算し、ダイナミックレンジを揃える。スケールファクタの計算は、レイヤ1では各帯域12サンプルごと、すなわち全体で384サンプルごとに、レイヤ2ではその3倍の1152サンプルを1ブロックとして384サンプルごとに行われる。このため、レイヤ2では解像度が増し、符号化品質が向上する。しかし、このままではレイヤ2のスケールファクタの数はレイヤ1の3倍になり、圧縮率の低下を招く。そこで、レイヤ2では3つのスケールファクタの組み合わせに応じて1つの新たな値(スケールファクタ選択情報)を割り当てて表現し、圧縮率低下を防ぐ。

【0113】図14は本発明の実施の形態8における音声再生装置のブロック図を示すものである。図14において、101はフレーム逆パッキング手段、102は逆量子化手段、103はデータ伸縮手段、104はサブバンド合成フィルタ手段、106はフレームカウント手段、12-1-1はエネルギー演算手段、12-1-2は伸縮頻度制御手段、12-1-3はフレーム選択手段、107はデータ伸縮制御手段である。

【0114】図15は、本発明の実施の形態8における、エネルギー演算手段12-1-1がフレームのエネルギーを推定する過程を示すフローチャートである。以下に、その動作について説明する。

【0115】本実施の形態8は、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して速度変換処理を施す音声再生装置の例を示すものである。MPEG1オーディオレイヤ2のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【0116】図14において、入力されたMPEG1オーディオレイヤ2のビットストリームは、フレーム逆パッキング手段101によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。

【0117】ここで、スケールファクタインデックスは、再生時の波形倍率を示し、各チャンネル、各有効サブバンド、各ブロックごとに存在する。スケールファクタインデックスは0から62までの値を取り、0が最もエネルギーが大きく、62が最もエネルギーが小さい。ただしビット割当情報が0の場合はスケールファクタインデックスは存在しない。また、ビット割当情報は、エンコード時に割当てべきビット数に関連した値で、各チャンネル、各有効サブバンドごとに存在する。

【0118】既に、述べたことでもあるが、MPEG1オーディオレイヤ2におけるチャンネルは、右チャンネルと左チャンネルの2チャンネル存在しうる。また、MPEG1オーディオレイヤ2におけるサブバンドは、全帯域を32等分割したものであり、周波数の低い順に、第0サブバンド、第1サブバンド、第2サブバンドから第31サブバンドまで存在する。

【0119】ここで、サブバンドについては、サンプリング周波数が32kHzの場合、0～16000Hzの帯域を32等分割するため、一つのサブバンドは500Hzの幅を持つ。ただし、32個のサブバンドのうち有効なサブバンド数が制限される。例えば192kbpsステレオの場合、0～31の32個のサブバンドのうち、0～29までの30個のサブバンドを有効サブバンドとするため、第30、第31サブバンドのビット割当情報や、スケールファクタインデックスは存在しない。この時、周波数帯域は0～15000Hzとなる(16000÷32×30=15000より)。

【0120】また、MPEG1オーディオレイヤ2におけるブロックとは、フレームを時間領域で3等分割した領域であり、時間順に第0ブロック、第1ブロック、第2ブロックまで存在する。サンプリング周波数が32kHzの場合、1ブロック長=12msである。1フレーム長は、サンプリング周波数が32kHzの場合36m

sである。

【0121】エネルギー演算手段12-1-1は、第0ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L0$ と、第0ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R0$ とを用いて、フレームサイクル内の各フレームナンバ $frm$ に対するエネルギー推定値 $e[frm]$ を求める。より詳しくは、スケールファクタインデックスの小さいフレームほどエネルギーは大きいもので、上記 $scf\_L0$ と $scf\_R0$ のうちどちらか小さい方の値を用いて、上記エネルギー推定値 $e[frm]$ を求める。

【0122】 $scf\_L0$ と $scf\_R0$ の一方が存在しない時は、エネルギー演算手段12-1-1は、存在するもう一方の値を用いて、エネルギー推定値 $e[frm]$ を求める。 $scf\_L0$ と $scf\_R0$ の両方が存在しない時は、エネルギー演算手段12-1-1は、速度変換フレーム選択候補の優先順位を最低にすることを意味する所定値を、エネルギー推定値 $e[frm]$ に代入する。

【0123】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数と、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。例えば0.9倍速の時、9フレームのうち2フレームを速度変換を施すフレームとして選択する。つまりフレームサイクル数は9であり、フレームナンバ $frm$ は0から8を変動する。フレーム選択手段12-1-3は、エネルギー演算手段12-1-1が出力するフレームサイクル中の全フレームに対するエネルギー推定値 $e[frm]$ の小さいフレームから順に、伸縮処理を行うフレームを選択する。 $e[frm]$ の小さいフレームを優先的に選択すれば、エネルギーの小さい音の部分が速度変換処理されることになる。

【0124】なお、第0ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L0$ と、第1ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L1$ と、第2ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L2$ と、第0ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R0$ と、第1ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R1$ と、第2ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R2$ とのうちの最小値を用いて、エネルギー推定値 $e[frm]$ を求めるようにしてもよい。

【0125】以上のように、本実施の形態8によれば、エネルギー演算手段12-1-1は、再生時の波形倍率を示すスケールファクタインデックスの値をもとに、音声信号のエネルギーを推定するようにし、その結果に応じて速度変換を施すフレームを選択するようにしたので、MP

EGデコード後のPCMデータのエネルギー演算が不要となり、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して、速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で速度変換処理を実現することができるものである。

【0126】(実施の形態9)以下、本発明の実施の形態9について、図面を参照しながら説明する。図16は、本発明の実施の形態9における音声再生装置のブロック図を示すものである。図16において、101はフレーム逆パッキング手段、102は逆量子化手段、103はデータ伸縮手段、104はサブバンド合成フィルタ手段、106はフレームカウント手段、13-1-1は定常性演算手段、12-1-2は伸縮頻度制御手段、13-1-3はフレーム選択手段、107はデータ伸縮制御手段である。表4は、本発明の実施の形態9において定常性演算手段13-1-1が出力する、定常性検出による速度変換フレーム選択優先順位である。以下に、その動作について説明する。

【0127】

【表4】

| ord [frm] | scfsi_L0 | scfsi_R0 |
|-----------|----------|----------|
| 1         | 2        | 2        |
| 2         | 1        | 2        |
| 2         | 2        | 1        |
| 2         | 2        | 3        |
| 2         | 3        | 2        |
| 3         | 1        | 1        |
| 3         | 1        | 3        |
| 3         | 3        | 1        |
| 3         | 3        | 3        |
| 4         | 0        | 2        |
| 4         | 2        | 0        |
| 5         | 0        | 1        |
| 5         | 1        | 0        |
| 5         | 0        | 3        |
| 5         | 3        | 0        |
| 6         | 0        | 0        |

【0128】本実施の形態9は、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して速度変換処理を施す音声再生装置の例を示すものである。MPEG1オーディオレイヤ2のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【0129】図16において、入力されたMPEG1オーディオレイヤ2のビットストリームは、フレーム逆パッキング手段101によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。スケールファクタ選択情報は、波形定常性を示すものであり、各チャンネル、各有効サブバンドごとに存在している。スケールファクタ選択情報は、0、1、2、3の値を取りうる。スケールファクタ選択情報が0のとき最も定常性が低く、2のとき最も定常性が高いものと見なす。スケールファクタ

選択情報が1および3のとき定常性は同等であると見なす。

【0130】定常性演算手段13-1-1は、第0サブバンドの左チャンネルのスケールファクタ選択情報 $scfsi\_L0$ と、第0サブバンドの右チャンネルのスケールファクタ選択情報 $scfsi\_R0$ とを用いて、フレームサイクル内の各フレームナンバ $f_{rm}$ に対する速度変換フレーム選択優先順位 $ord[f_{rm}]$ を求める。定常性演算手段13-1-1は、フレームサイクル内の全フレームの $ord[f_{rm}]$ を、表4に示す規則に従って求める。 $scfsi\_L0$ と $scfsi\_R0$ のどちらか一つまたは両方が存在しないときは、定常性演算手段13-1-1は、速度変換フレーム選択候補の優先順位を最低であることを意味する所定値を、速度変換フレーム選択優先順位 $ord[f_{rm}]$ に代入する。

【0131】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数とそのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。フレーム選択手段13-1-3は、定常性演算手段13-1-1が出力するフレームサイクル中の全フレームに対する速度変換フレーム選択優先順位 $ord[f_{rm}]$ の高いフレームから順に、伸縮処理を行うフレームを選択する。

【0132】以上のように、本実施の形態9によれば、定常性演算手段13-1-1は、波形定常性を示すスケールファクタ選択情報の値をもとに、音声信号の定常性を推定することにより、MPEGデコード後のPCMデータの定常性演算が不要となり、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で速度変換処理実現することができるものである。

【0133】このような、本実施の形態9では、速度変換による音質劣化の少ない定常性の高いフレームを選択して速度変換するということに特徴があり、このように、話速変換ができるので、語学学習に適しているものであり、また、定常性演算処理が不要となるため、演算量を削減できる、という特徴をも有するものである。

【0134】(実施の形態10)以下、本発明の実施の形態10について、図面を参照しながら説明する。図17は、本発明の実施の形態10における音声再生装置のブロック図を示すものである。図17において、101はフレーム逆パッキング手段、102は逆量子化手段、103はデータ伸縮手段、104はサブバンド合成フィルタ手段、106はフレームカウント手段、14-1-1はエネルギー変化度合演算手段、12-1-2は伸縮頻度制御手段、14-1-3はフレーム選択手段、107はデータ伸縮制御手段である。以下に、その動作について説明する。

【0135】本実施の形態10は、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して速度変換処理を施す音声再生装置の例を

示すものである。MPEG1オーディオレイヤ2のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【0136】図17において、入力されたMPEG1オーディオレイヤ2のビットストリームは、フレーム逆パッキング手段101によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。

【0137】エネルギー変化度合演算手段14-1-1は、第0ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L0$ と、第1ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L1$ と、第2ブロックの第0サブバンドの左チャンネルのスケールファクタインデックス $scf\_L2$ と、第0ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R0$ と、第1ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R1$ と、第2ブロックの第0サブバンドの右チャンネルのスケールファクタインデックス $scf\_R2$ とを用いて、フレームサイクル内と、フレームサイクルの前後1フレーム、の各フレームナンバ $f_{rm}$ に対する各チャンネルの各ブロックのエネルギー推定値 $e[ch][blk][f_{rm}]$ を求める。フレームサイクル内とフレームサイクルの前後1フレームとは、例えばフレームサイクル数が9の場合、9フレームの前後1フレームということで、11フレームとなる。

【0138】即ち、各ブロックの第0サブバンドの各チャンネルのスケールファクタインデックスに対応する、各フレームナンバの各チャンネルの各ブロックのエネルギー推定値 $e[ch][blk][f_{rm}]$ を、フレームサイクル内とフレームサイクルの前後1フレームについて求める。スケールファクタインデックスの小さいブロックほどエネルギーは大きい。

【0139】また、スケールファクタインデックスが存在しないとき、エネルギーは0である。即ち、 $scf\_L0$ が存在しないフレームの場合、 $e[0][0][f_{rm}]=0$ とする。 $scf\_L1$ が存在しないフレームの場合、 $e[0][1][f_{rm}]=0$ とする。 $scf\_L2$ が存在しないフレームの場合、 $e[0][2][f_{rm}]=0$ とする。 $scf\_R0$ が存在しないフレームの場合、 $e[1][0][f_{rm}]=0$ とする。 $scf\_R1$ が存在しないフレームの場合、 $e[1][1][f_{rm}]=0$ とする。 $scf\_R2$ が存在しないフレームの場合、 $e[1][2][f_{rm}]=0$ とする。

【0140】次に、フレームサイクル内の各フレームナンバ $f_{rm}$ に対するエネルギー推定値 $e[ch][blk]$

$k$ ][ $f_{rm}$ ]のブロック内の最大値 $e_{max}[ch][f_{rm}]$ を、全フレームサイクルについて求める。フレームサイクルの前後1フレームの $e_{max}[ch][f_{rm}]$ は、求めなくてよい。

【0141】次に、フレームサイクル内の各フレームナンバ $f_{rm}$ に対して、エネルギー推定値 $e[0][2][f_{rm}-1]-e_{max}[0][f_{rm}]$ と、エネルギー推定値 $e[1][2][f_{rm}-1]-e_{max}[1][f_{rm}]$ と、エネルギー推定値 $e[0][0][f_{rm}+1]-e_{max}[0][f_{rm}]$ と、エネルギー推定値 $e[1][0][f_{rm}+1]-e_{max}[1][f_{rm}]$ の4個の値を求め、4個の値のうちの最大値を、速度変換フレーム選択優先度 $p[f_{rm}]$ に代入する。

【0142】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数と、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。フレーム選択手段14-1-3は、エネルギー変化度合演算手段14-1-1が出力するフレームサイクル中の全フレームに対する速度変換フレーム選択優先度 $p[f_{rm}]$ の大きいフレームから順に、伸縮処理を行うフレームを選択する。速度変換フレーム選択優先度 $p[f_{rm}]$ の大きいフレームほど、非同時マスキングでマスキングされやすいので、速度変換による音質劣化が知覚されにくいことが特徴となる。非同時マスキングについては、B、C、J、ムーア著、誠信書房発行、聴覚心理学概論に詳しく記述されている。

【0143】以上のように、本実施の形態10によれば、エネルギー変化度合演算手段14-1-1は、再生時の波形倍率を示すスケールファクタインデックスの値をもとに、音声信号のエネルギー変化度合を推定し、速度変換フレーム選択優先度 $p[f_{rm}]$ の大きいフレームを優先的に速度変換するようにしたので、MPEGデコード後のPCMデータのエネルギー変化度合演算が不要となり、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で実現できることが特徴である。また、この方法は、話速変換ができるので、語学学習に適した音声処理を行うことができる。

【0144】(実施の形態11)以下、本発明の実施の形態11について、図面を参照しながら説明する。図18は、本発明の実施の形態11における音声再生装置のブロック図を示すものである。図18において、101はフレーム逆パッキング手段、102は逆量子化手段、103はデータ伸縮手段、104はサブバンド合成フィルタ手段、106はフレームカウント手段、12-1-1はエネルギー演算手段、13-1-1は定常性演算手段、14-1-1はエネルギー変化度合演算手段、12-1-2は伸縮頻度制御手段、15-1-3はフレーム選択手段、107はデータ伸縮制

御手段である。以下に、その動作について説明する。

【0145】本実施の形態11は、MPEG1オーディオレイヤ2のビットストリームをデコードする際の間データに対して速度変換処理を施す音声再生装置の例を示すものである。MPEG1オーディオレイヤ2のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【0146】図18において、入力されたMPEG1オーディオレイヤ2のビットストリームは、フレーム逆パッキング手段101によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。

【0147】エネルギー演算手段12-1-1は、本発明の実施の形態8に記述した方法で、フレームサイクル内の各フレームナンバ $f_{rm}$ に対するエネルギー推定値 $e[f_{rm}]$ を求める。

【0148】定常性演算手段13-1-1は、本発明の実施の形態9に記述した方法で、フレームサイクル内の各フレームナンバ $f_{rm}$ に対する速度変換フレーム選択優先順位 $ord[f_{rm}]$ を求める。

【0149】エネルギー変化度合演算手段14-1-1は、本発明の実施の形態10に記述した方法で、フレームサイクル内の各フレームナンバ $f_{rm}$ に対する速度変換フレーム選択優先度 $p[f_{rm}]$ を求める。

【0150】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数と、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。自然性の劣化を少なく、効率的に聴取したい場合、フレーム選択手段15-1-3は、エネルギー演算手段12-1-1が出力するフレームサイクル中の全フレームに対するエネルギー推定値 $e[f_{rm}]$ の小さいフレームから順に、伸縮処理を行うフレームを選択する。了解性を高め、聴き取りやすい音声を得たい場合、フレーム選択手段15-1-3は、定常性演算手段13-1-1が出力するフレームサイクル中の全フレームに対する速度変換フレーム選択優先順位 $ord[f_{rm}]$ の高いフレームから順に、伸縮処理を行うフレームを選択する。このときフレームサイクル内の速度変換フレーム選択優先順位 $ord[f_{rm}]$ の値が同一で優先順位がつけられない場合は、エネルギー変化度合演算手段14-1-1が出力する速度変換フレーム選択優先度 $p[f_{rm}]$ を用いて、その $p[f_{rm}]$ の大きいフレームを優先的に選択するようにして速度変換フレーム選択優先順位 $ord[f_{rm}]$ の値が同一なフレームに対して細分化した優先順位をつける。

【0151】以上のように、本実施の形態11によれば、エネルギー演算手段12-1-1と、定常性演算手段13-1-1と、エネルギー変化度合演算手段14-1-1は、再生時の波形倍率を示すスケールファクタインデックスと、スケ

ールファクタ選択情報の値をもとに、音声信号のエネルギーと、定常性と、エネルギー変化度合を推定し、自然性重視の場合、 $e[frm]$ の小さいフレームを選択し、了解性重視の場合、 $ord[frm]$ の小さいフレームを選択し、 $ord[frm]$ が同一の値の場合、 $p[frm]$ の大きいフレームを優先的に選択するようにしたので、MPEGデコード後のPCMデータのエネルギーと、定常性と、エネルギー変化度合の演算が不要となり、MPEG1オーディオレイヤ2のビットストリームをデコードする際の中間データに対して速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で所望の話速変換処理を行うことができる効果を得られる。

【0152】なお、本発明の実施の形態4に記載されている音声らしさ演算手段4004が本実施の形態11に記載されていないのは、MPEG1オーディオレイヤ2のビットストリームに音声らしさを示す情報が含まれていないためである。

【0153】[参考文献]

- 1) 鈴木, 三崎, “高品質速度変換方式のDSPによる実現”, 信学技報, SP90-34(1990)
- 2) 比企他, “連続音声の音韻区分の持続時間の性質”, 信学誌, 第50巻5号, pp.849-856(1967)
- 3) 中山, 三, “日本人学習者に対する英語の語頭強調処理による受聴明瞭度の改善”, 音講論集, 1-8-21(1998.3)
- 4) 細井, 目方他, “補聴効果評価のための67-S早口語音聴力検査”, Audiology Japan, vol.36.No.5, pp.299-300(1993)
- 5) B.C.J.Moore 著 (大串健吾監訳), “聴覚心理学概論” 誠信書房 (非同期マスキングに関しても参照)

【0154】

【発明の効果】請求項1にかかる音声再生装置によれば、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して時間軸変換処理を施すデータ伸縮手段と、与えられた速度比に応じて各フレームに対する伸縮処理のシーケンスが定められたフレームシーケンステーブルと、上記入力音声信号のフレーム数をカウントするフレームカウント手段と、上記フレームカウント手段からのカウント値をもとに上記フレームシーケンステーブルを参照し上記データ伸縮手段に対して当該フレームを時間軸圧縮/時間軸伸長/時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記データ伸縮手段は、上記データ伸縮制御手段からの指示信号に基づいて音声信号の時間軸変換処理を行うものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮/伸長処理を基本とした簡素な構成によって、所望の速度比(再生速度)にて高品質な速度変換処理を実現する音声再生装置を提供することができる効果がある。

【0155】また、請求項2にかかる音声再生装置によれば、請求項1に記載の音声再生装置において、データ伸縮手段は、入力された音声信号の各フレームを少なくとも2つのセグメントに分割して各セグメントを構成する波形データを重み付け加算するクロスフェード手段を具備するものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮/伸長処理を基本とした簡素な構成によって、所望の速度比(再生速度)にて高品質な速度変換処理を実現する音声再生装置を提供することができる効果がある。

【0156】また、請求項3にかかる音声再生装置によれば、請求項1または2に記載の音声再生装置において、データ伸縮手段は、フレームに対して所定比率の時間軸圧縮/伸長を行い、データ伸縮制御手段は、時間軸圧縮/伸長を施すフレームと、時間軸変換なしで出力するフレームとの出現頻度を制御することにより、上記与えられた速度比で音声再生するものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮/伸長処理を基本とした簡素な構成によって、所望の速度比(再生速度)にて高品質な速度変換処理を実現する音声再生装置を提供することができる効果がある。

【0157】また、請求項4にかかる音声再生装置によれば、請求項3に記載の音声再生装置において、データ伸縮手段は、フレームに対して所定比率の時間軸圧縮/伸長を行い、フレームシーケンステーブルには時間軸圧縮/伸長のシーケンスが繰り返されるフレームサイクル内において時間軸圧縮/伸長を施すフレームを可能な限り均等に配置するものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮/伸長処理を基本とした簡素な構成によって、所望の速度比(再生速度)にて高品質な速度変換処理を実現する音声再生装置を提供することができる効果がある。

【0158】また、請求項5にかかる音声再生装置によれば、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と、与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号のエネルギーを演算するエネルギー演算手段と、上記エネルギー演算手段の出力と、伸縮頻度制御手段の出力とをもとに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮/時間軸伸長/時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、エネルギーの小さいフレームを優先的に選択するものとしたので、エネルギーの小さいフレームでの時間軸伸縮は処理劣化が検知され難いため、高品質な速度変換処理音声を得ることができる。

【0159】また、請求項6にかかる音声再生装置によ

れば、入力される音声信号を、フレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、当該フレームに含まれる人間の音声を含む可能性を演算する音声らしさ演算手段と、上記音声らしさ演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、音声らしさの小さいフレームを優先的に選択するものとしたので、音声情報が含まれないと推定されるフレームでの時間軸伸縮は処理劣化が検知され難いため、高品質な速度変換処理音声を得ることができる。

【0160】また、請求項7にかかる音声再生装置によれば、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号の定常性を演算する定常性演算手段と、上記定常性演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、定常性の高いフレームを優先的に選択するものとしたので、定常性の高いフレームでは重み付け加算法による劣化が検知され難いため、高品質な速度変換処理音声を得ることができる。

【0161】また、請求項8にかかる音声再生装置によれば、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号のエネルギーの変化度合いを演算するエネルギー変化度合演算手段と、上記エネルギー変化度合演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、エネルギー変化度合をともに、処理劣化が継続マスクング効果によって検知しにくいフレームを優先的に選択する

ものとしたので、継続マスクング効果により処理劣化が検知され難いため、高品質な速度変換処理音声を得ることができる。

【0162】また、請求項9にかかる音声再生装置によれば、入力される音声信号をフレーム単位で復号する音声復号化手段と、上記復号されたフレーム内のデータに対して、時間軸変換処理を施すデータ伸縮手段と与えられた速度比に応じて、フレームサイクルと、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する伸縮頻度制御手段と、フレーム内における音声信号のエネルギーを演算するエネルギー演算手段、あるいは、当該フレームに含まれる人間の音声を含む可能性を演算する音声らしさ演算手段、あるいは、フレーム内における音声信号の定常性を演算する定常性演算手段、あるいは、フレーム内における音声信号のエネルギー変化度合を演算するエネルギー変化度合演算手段のうち2つ以上の演算手段を備え、上記複数の演算手段の出力と、伸縮頻度制御手段の出力とをともに、伸縮すべきフレームを選択するフレーム選択手段と、上記データ伸縮手段に対して、当該フレームを時間軸圧縮／時間軸伸長／時間軸変換なしのいずれで処理するかを指示するデータ伸縮制御手段とを具備し、上記フレーム選択手段は、上記複数の演算手段の出力を総合的に判断して、選択すべきフレームを決定するものとしたので、自然性を重視した再生方法や、了解性を重視した再生方法などを使用者が選択することができ、目的に応じてそれぞれ高品質な速度変換処理音声を得ることができる。

【0163】また、請求項10にかかる音声再生装置によれば、請求項1ないし9のいずれかに記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号を行うものとしたので上記請求項1ないし9の各項の効果を得られるものを実現できる効果がある。

【0164】また、請求項11にかかる音声再生装置によれば、請求項1ないし10のいずれかに記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEGオーディオ符号化方式にて符号化されたデータを復号するものとしたので、かかるMPEGオーディオ符号化方式にて符号化されたデータにて処理劣化の少ない速度変換処理を行なうことができる。

【0165】また、請求項12にかかる音声再生装置によれば、請求項5に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、エネルギー演算手段は、再生時の波形倍率を示すスケールファクタインデックスをともに音声信号のエネルギーを推定するものとしたので、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータに対して、処理劣化の少ない測度変換処理を行なうことができる。

【0166】また、請求項13にかかる音声再生装置によれば、請求項7に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、定常性演算手段は、波形定常性を示すスケールファクタ選択情報をもとに音声信号の定常性を推定するものとしたので、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータに対して、処理劣化の少ない速度変換処理を行なうことができる。

【0167】また、請求項14にかかる音声再生装置によれば、請求項8に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、エネルギー変換度合演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギー変化度合を推定するものとしたので、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータに対して、処理劣化の少ない速度変換処理を行なうことができる。

【0168】また、請求項15にかかる音声再生装置によれば、請求項9に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータを復号し、エネルギー演算手段、定常性演算手段、エネルギー変換度合演算手段のうち2つ以上の演算手段を備え、エネルギー演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギーを推定し、定常性演算手段は、波形定常性を示すスケールファクタ選択情報をもとに音声信号の定常性を推定し、エネルギー変換度合演算手段は、再生時の波形倍率を示すスケールファクタインデックスをもとに音声信号のエネルギー変化度合を推定するものとしたので、MPEG1オーディオレイヤ2符号化方式にて符号化されたデータに対して、処理劣化の少ない速度変換処理を行なうことができる。

【0169】また、請求項16にかかる音声再生装置によれば、請求項1ないし15のいずれかに記載の音声再生装置において、データ伸縮手段は、フレームを構成する各セグメントの相関を求め、各セグメント間の相関が高くなる位置を演算により求めて、当該位置にセグメントの波形データをシフトさせるためのシフト量を送出する相関演算手段を具備し、クロスフェード手段は、上記シフト量をもとにセグメントの波形データをシフトさせて各セグメントデータの重み付け加算を行い、次回に時間軸圧縮／伸長されるフレームは、前回時間軸圧縮／伸長されたフレームに関する上記シフト量を考慮のうえ、セグメントデータがシフトされ重み付け加算が施されるものとしたので、フレームを構成するセグメント間の相関が高くなる位置に波形データをシフトさせて相関演算を行う、かつ各時間軸圧縮／伸長の処理において上記シ

フト量を考慮した処理を行なうことができる。

【0170】また、請求項17にかかる音声再生装置によれば、請求項1ないし16のいずれかに記載の音声再生装置において、データ伸縮手段は、フレームを構成する各セグメントの相関を求める相関演算手段を具備し、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号が成され、上記相関演算手段は、音声信号のピッチ周波数を包含する帯域のデータを用いて上記各セグメントの相関を求めるものとしたので、上記各セグメントの相関を求める演算において音声信号のピッチ周波数を包含する帯域のデータを求めることにより、音声信号の処理劣化の少ない速度変換処理を行なうことができる。

【0171】また、請求項18にかかる音声再生装置によれば、請求項16に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号が成され、上記相関演算手段は、相関値の演算を各分割帯域毎に行い、最大の相関値を持つ帯域のシフト量を用いて重み付け加算を行うものとしたので、かかる相関演算により、音声信号の処理劣化の少ない速度変換処理を行なうことができる。

【0172】また、請求項19にかかる音声再生装置によれば、請求項16に記載の音声再生装置において、フレーム単位で復号を行う音声復号化手段は、音声信号を複数の帯域に分割し当該帯域分割された各帯域毎に復号が成され、上記相関演算手段は、相関値の演算を各分割帯域のうちの最も平均エネルギーの大きい帯域に対して行うものとしたので、かかる相関演算により、音声信号の処理劣化の少ない速度変換処理を行なうことができる。

【0173】また、請求項20にかかる音声再生装置によれば、請求項1ないし4のいずれか、または請求項16ないし19のいずれかに記載の音声再生装置において、フレームシーケンステーブルは、1つの速度比に対してパターンの異なる複数のシーケンステーブルを備え、データ伸縮手段は、予め伸縮フレームにおける各セグメントの相関値の平均値を各シーケンステーブル毎に求め、最も相関値の平均が高いシーケンステーブルを参照して処理を行うものとしたので、かかるデータ伸縮処理により、音声信号の処理劣化の少ない速度変換処理を行なうことができる。

【図面の簡単な説明】

【図1】本発明の実施の形態1による音声再生装置の全体ブロック図。

【図2】本発明の実施の形態1におけるデータ伸縮手段の構成図。

【図3】本発明の実施の形態1におけるデータ伸縮手段における一定値の時間軸圧縮／伸長の様子を示す 模式図。

【図4】本発明の実施の形態1における伸縮シーケンスの模式図。

【図5】本発明の実施の形態2におけるデータ伸縮手段の構成図。

【図6】本発明の実施の形態2におけるデータ圧縮の模式図。

【図7】本発明の実施の形態2におけるデータ圧縮の補正を行う場合のの模式図。

【図8】本発明の実施の形態2における他の例の伸縮シーケンスの模式図。

【図9】本発明の実施の形態3による音声再生装置の全体ブロック図。

【図10】本発明の実施の形態4による音声再生装置のブロック図。

【図11】本発明の実施の形態5による音声再生装置のブロック図。

【図12】本発明の実施の形態6による音声再生装置のブロック図。

【図13】本発明の実施の形態7による音声再生装置のブロック図。

【図14】本発明の実施の形態8による音声再生装置のブロック図。

【図15】本発明の実施の形態8における、エネルギー演算手段12-1-1がフレームのエネルギーを推定する過程を示すフローチャートである。

【図16】本発明の実施の形態9による音声再生装置のブロック図。

【図17】本発明の実施の形態10による音声再生装置のブロック図。

【図18】本発明の実施の形態11による音声再生装置のブロック図。

【図19】従来の音声再生装置のブロック図。

【図20】従来の他の例の音声装置のブロック図。

【図21】音声信号の主要ピッチ成分が含まれる周波数帯域について、その1フレーム分の時間軸波形を表した図。

【図22】図21に示した1フレームの信号を、その前半の信号部分と、後半の信号部分との2セグメントに分割して上下に並べた図。

【図23】図22における2セグメント間の相関関数を求めた値を示したグラフ。

【図24】相関関数が最大となる時刻に後半の信号成分であるセグメントをずらせた様子を定性的に示した図。

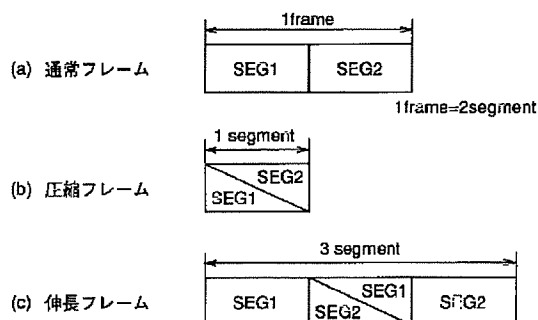
【図25】2セグメント間を $T_c$ 時間オーバーラップさせてクロスフェード処理する様子を示した図。

【図26】MPEG1 オーディオレイヤ2の構成を示すブロック図。

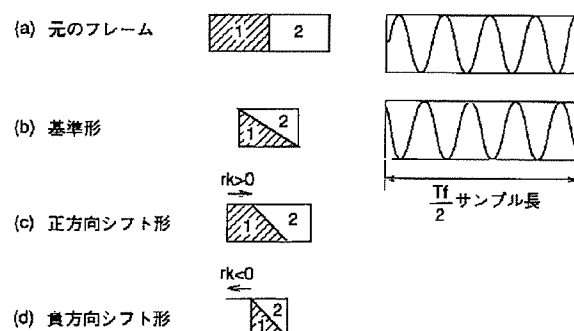
#### 【符号の説明】

- 101 フレーム逆バッキング手段
- 102 逆量子化手段
- 103 データ伸縮手段
- 104 サブバンド合成フィルタ手段
- 105 選択手段
- 106 フレームカウント手段
- 107 データ伸縮制御手段
- 108 フレームシーケンステーブル
- 201 バッファメモリ
- 202 クロスフェード手段
- 203 データ選択手段
- 301 相関演算手段
- 302 位相制御記憶手段
- 3001 フレーム複号化手段
- 3002 データ伸縮手段
- 3003 伸縮頻度制御手段
- 3004 エネルギー演算手段
- 3005 フレーム選択手段
- 3006 データ伸縮制御手段
- 4004 音声らしさ演算手段
- 4005 フレーム選択手段
- 5004 定常性演算手段
- 5005 フレーム選択手段
- 6004 エネルギー変化度合演算手段
- 6005 フレーム選択手段
- 7005 フレーム選択手段

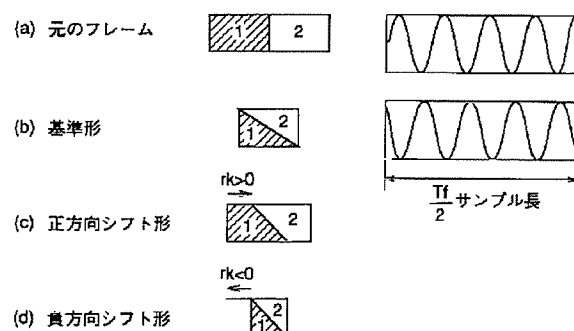
【図3】



【図6】

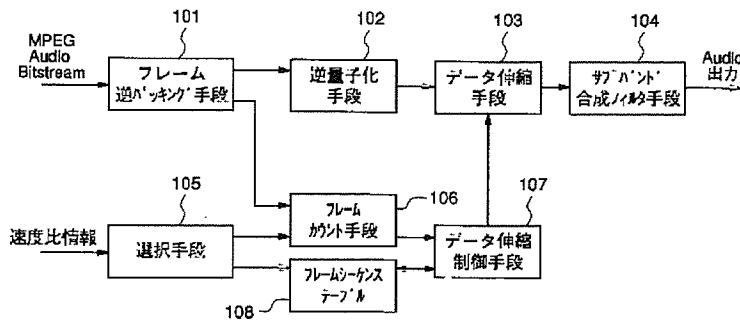


【図22】

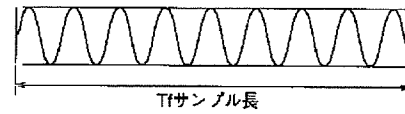




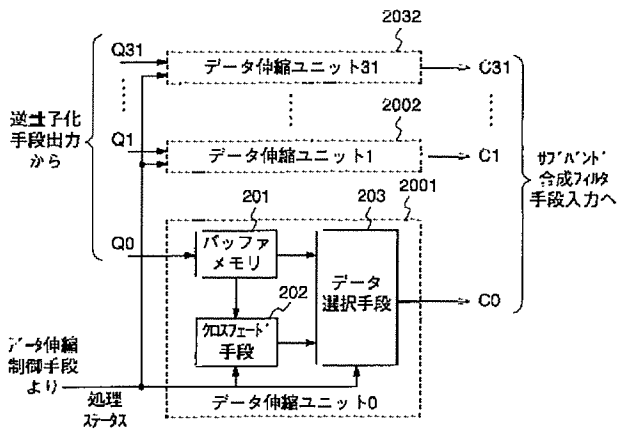
【図1】



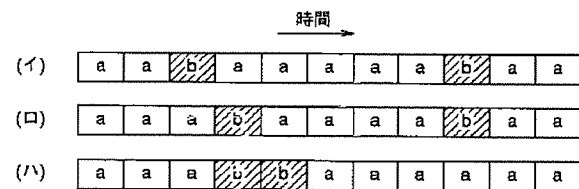
【図21】



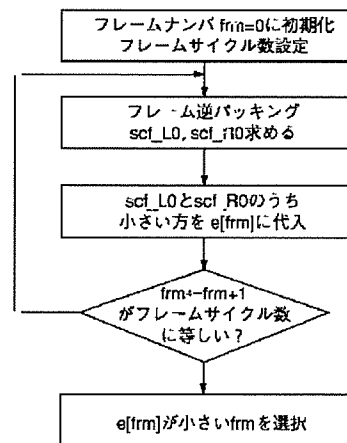
【図2】



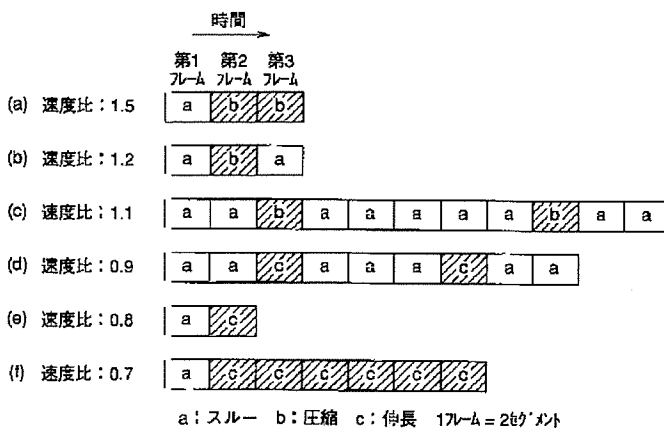
【図8】



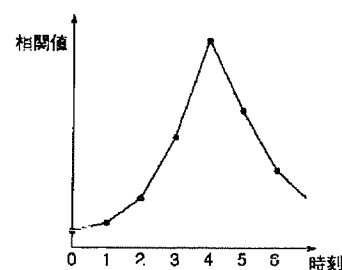
【図15】



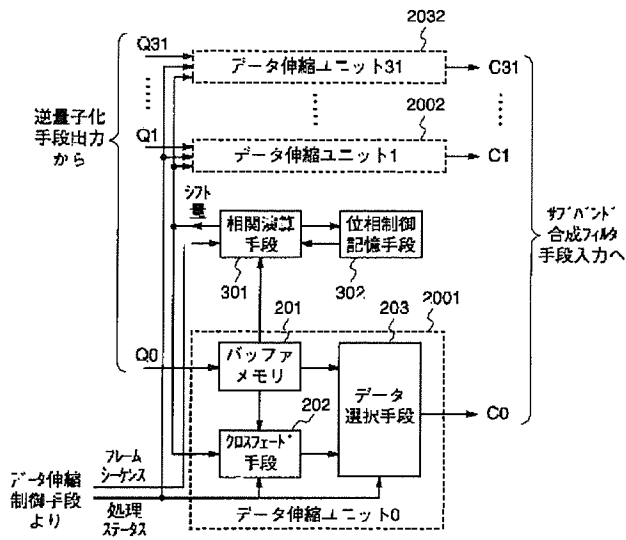
【図4】



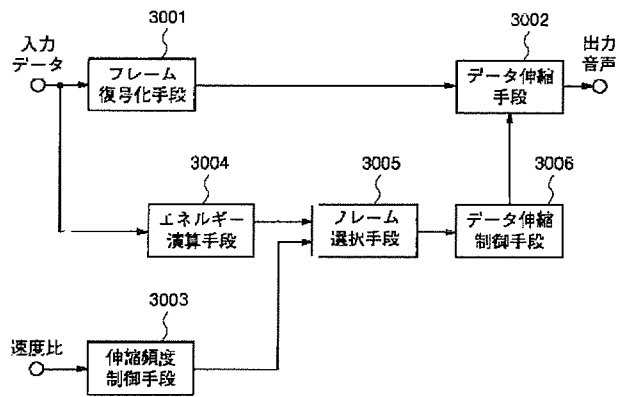
【図23】



【図5】

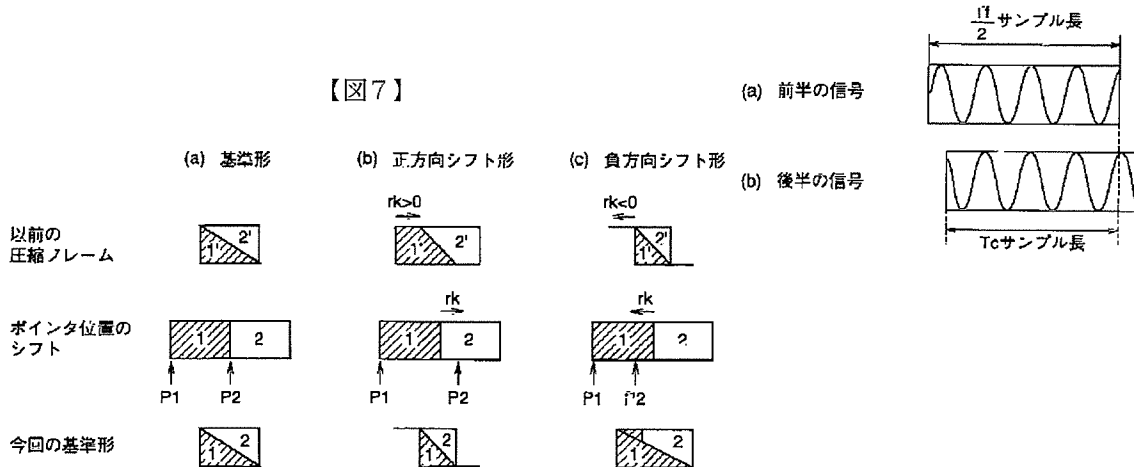


【図9】

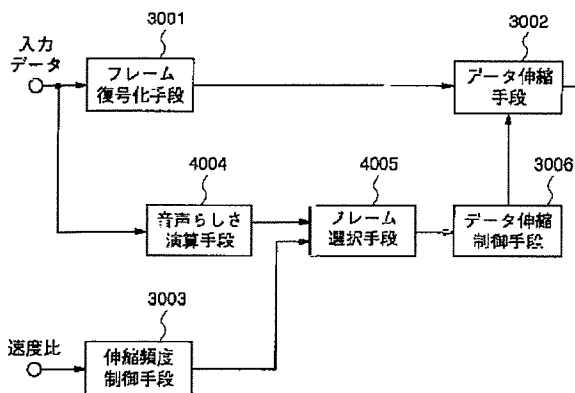


【図24】

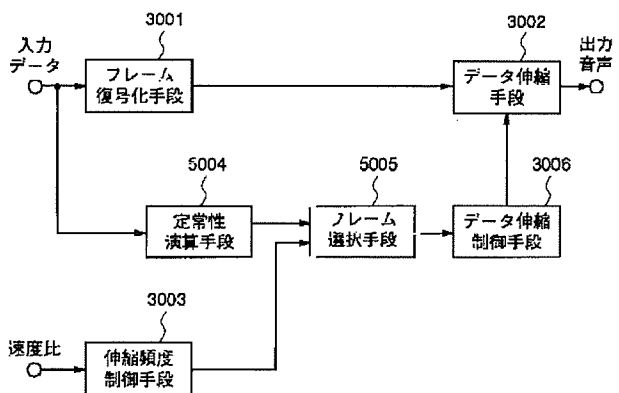
【図7】



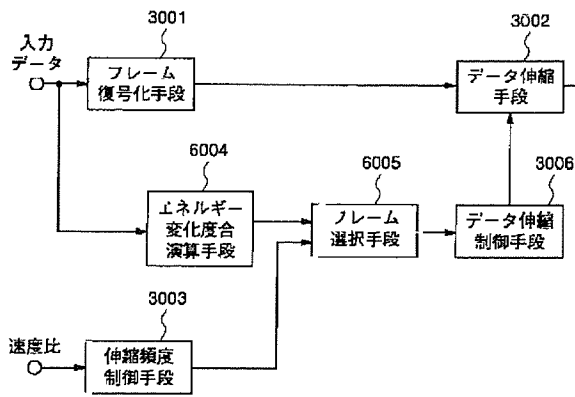
【図10】



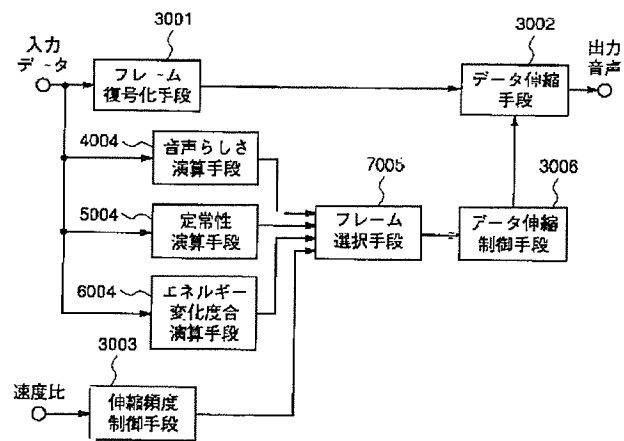
【図11】



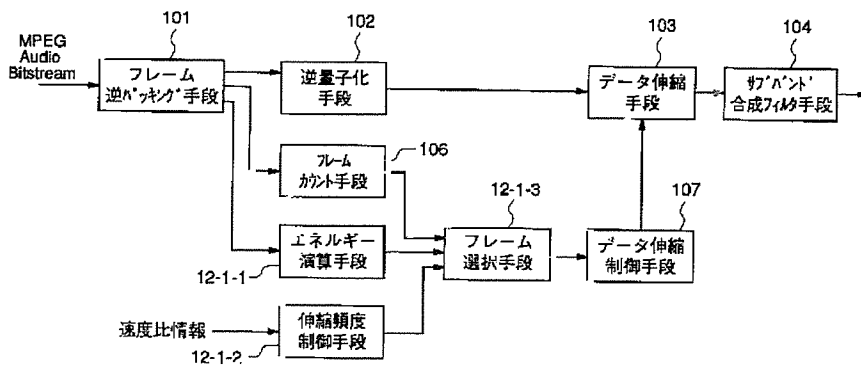
【図12】



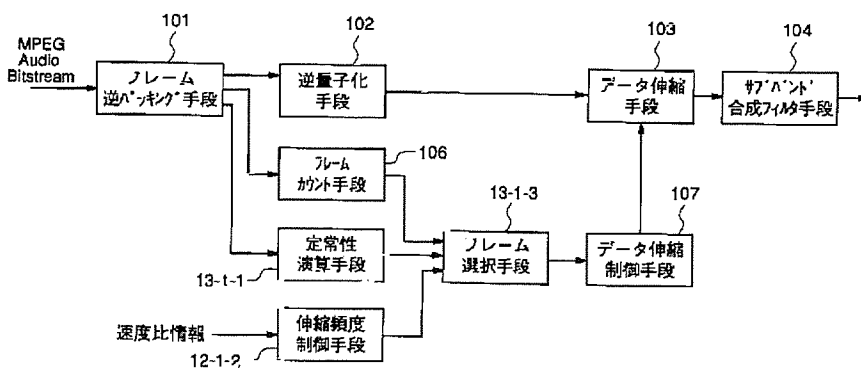
【図13】



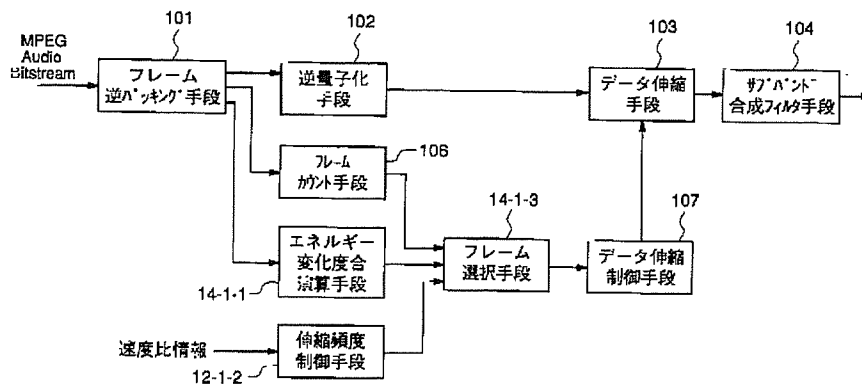
【図14】



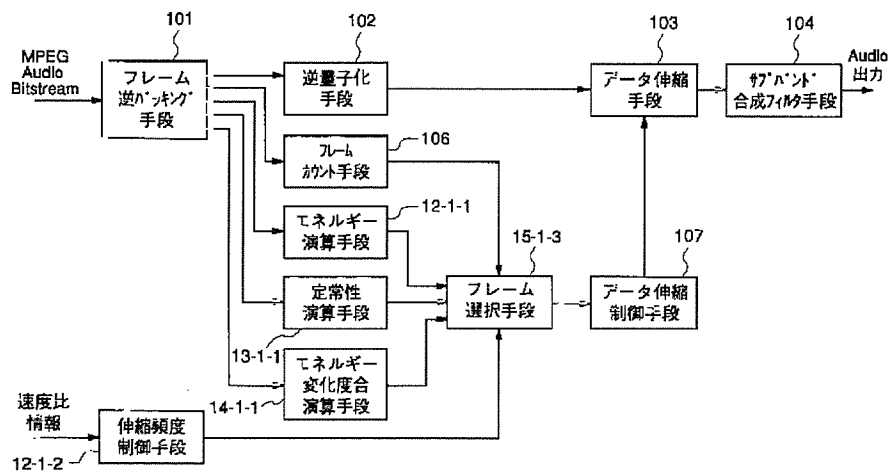
【図16】



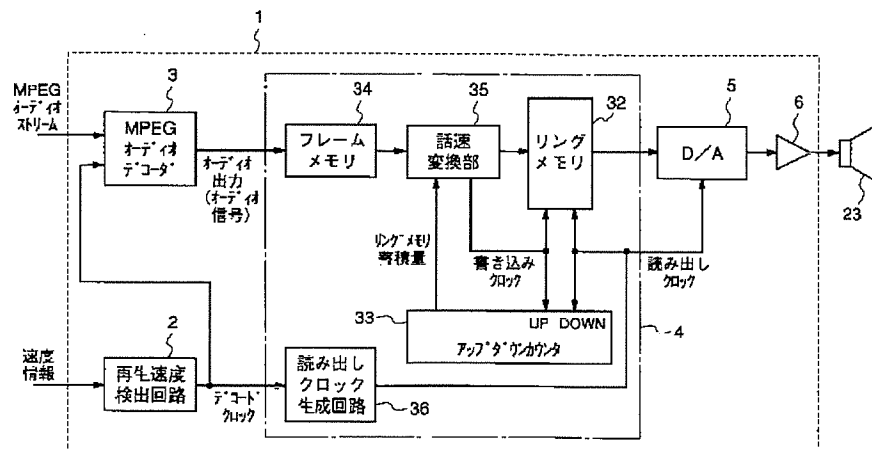
【図17】



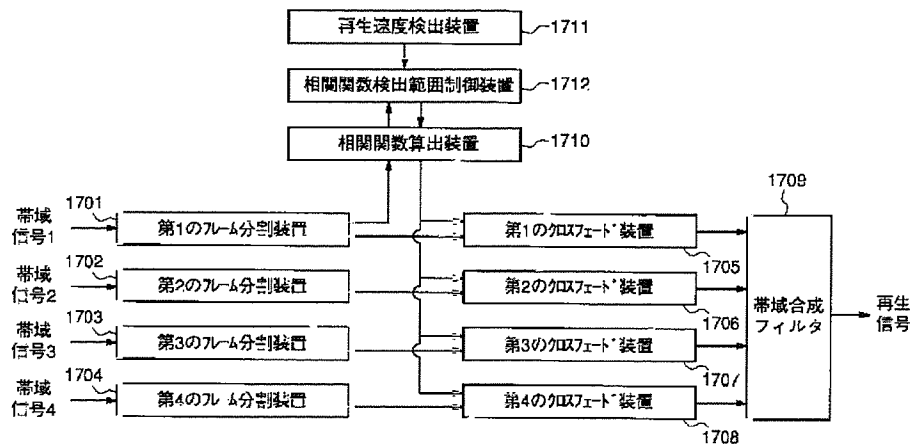
【図18】



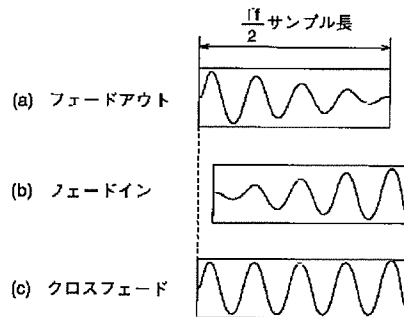
【図19】



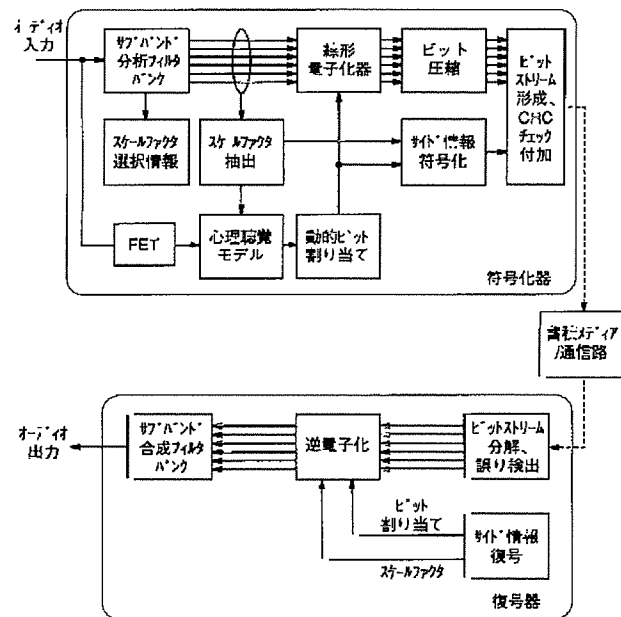
【図20】



【図25】



【図26】



フロントページの続き

(72)発明者 松本 美治男  
 大阪府門真市大字門真1006番地 松下電器  
 産業株式会社内